NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

GESTÃO DA QUALIDADE DOS DADOS EM CONTEXTO DE DADOS ABERTOS

Caso de Estudo de Lisboa

Rúben Gonçalo Teixeira Pagaime

Trabalho de Projeto apresentado como requisito parcial para obtenção do grau de Mestre em Gestão de Informação

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

GESTÃO DA QUALIDADE DOS DADOS EM CONTEXTO DE DADOS ABERTOS

CASO DE ESTUDO DE LISBOA

por

Rúben Gonçalo Teixeira Pagaime

Trabalho de Projeto apresentado como requisito parcial para a obtenção do grau de Mestre em Gestão de Informação, Especialização em Gestão do Conhecimento e Business Intelligence

Orientador/Coorientador: Professor Miguel de Castro Neto, PhD

Novembro 2018

AGRADECIMENTOS

Gostaria de agradecer ao meu orientador Dr. Miguel Neto pela disponibilidade demonstrada em assumir a orientação deste projeto, assim como pelo apoio prestado.

Gostaria também de agradecer à minha família e aos meus amigos pelo apoio demonstrado ao longo deste caminho, e pela compreensão demonstrada devido à minha falta de disponibilidade para estar presente em diversos eventos de convivência social.

Não posso deixar de agradecer à minha entidade empregadora e aos seus responsáveis pela compreensão e flexibilidade demonstrada, que facilitaram em diversas ocasiões a minha ausência no local de trabalho.

Gostaria também de agradecer aos meus colegas que estiveram envolvidos comigo nos mais diversos projetos do curso, nomeadamente a Ana Sousa, o Ricardo Apolinário e a Sónia Bernardes, pois foram fundamentais neste percurso.

Por fim, agradeço a todos os docentes da NovaIMS que estiveram presentes neste percurso académico e à escola pelas condições proporcionadas.

RESUMO

Nos últimos anos foi verificado um aumento considerável do entusiamo em relação à temática dos dados abertos. Dados Abertos são dados que estão disponíveis publicamente e que podem ser acedidos, utilizados e redistribuídos gratuitamente. No entanto, a reutilização destes dados para a criação de novos produtos e novos serviços tem como base a qualidade dos mesmos. A falta de qualidade dos dados abertos que são disponibilizados publicamente pode comprometer o objetivo da sua reutilização.

É neste contexto que foi desenvolvido este trabalho de projeto, que consiste no desenvolvimento de uma *framework* que tenha como resultado a indicação de diversos parâmetros da qualidade dos Dados Abertos que são disponibilizados no Portal de Dados Abertos da cidade de Lisboa.

A análise da literatura será iniciada com uma descrição dos conceitos de *Smart Cities*, da influência da '*Internet of Things'* (*IoT)* e do surgimento do conceito de Dados Abertos dentro do contexto das *Smart Cities*.

Será também efetuado um estudo da temática dos Dados Abertos, nomeadamente sobre a qualidade dos mesmos, a sua gestão e metodologias para a avaliação e melhoria dessa mesma qualidade, de modo a que possam ser retiradas métricas de avaliação para serem aplicadas sobre os dados do Portal de Dados Abertos da cidade de Lisboa.

Após a aplicação dessas métricas, serão apresentados os resultados, de modo a que exista uma boa ideia do estado atual da qualidade dos dados que são disponibilizados no Portal de Dados Abertos da cidade de Lisboa, dentro dos parâmetros estudados.

Com base neste estudo serão deixadas indicações e orientações para que se consigam resultados ainda melhores em futuras investigações sobre este tema, nomeadamente com o estudo de outros parâmetros que não serão investigados neste trabalho de Projeto.

PALAVRAS CHAVE

Smart Cities; Lisboa; Dados Abertos; Qualidade dos Dados; *Dashboard*

ABSTRACT

In recent years there has been a considerable increase in the enthusiasm for open data. Open Data is data that is publicly available and can be accessed, used and redistributed free of charge. However, the reuse of this data for the creation of new products and new services is based on their quality. The lack of quality of open data that is made publicly available may undermine the purpose of its reuse.

It is in this context that this project work was developed, which consists on the development of a framework that results in the indication of several parameters of the Open Data quality that are available in the Open Data Portal of the city of Lisbon.

The literature review will begin with a description of Smart Cities concepts, the 'Internet of Things' (IoT) influence and the emergence of the Open Data concept within Smart Cities context.

Will be carried out an Open Data study, the quality of the data, its management and methodologies for the evaluation and improvement of its quality, so that, evaluation metrics can be taken to be applied to the data of the Portal of Open Data of the city of Lisbon.

After applying these metrics, there will be a presentation of the results, so that we can see the current state of data quality that are available in the Open Data Portal of the city of Lisbon, within the studied parameters.

Based on this study, indications and guidelines will be left to obtain even better results in future research on this topic, namely with the study of other parameters that will not be investigated in this Project work.

KEYWORDS

Smart Cities; Lisbon; Open Data; Data Quality; Dashboard

ÍNDICE

ÍNDICE DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS E ABREVIATURAS

CML	Câmara Municipal de Lisboa
CKAN	Comprehensive Knowledge Archive Network
DSRM	Design Science Research Methodology
eOTD	ECCMA Open Technical Dictionary
ETL	Extract, Transform, Load
ETP	Extract, Transform, Publish
GPS	Global Position System
IoT	Internet of Things
ISO	International Organization for Standardization
IT	Information Technology
OKFN	Open Knowledge International
TIC	Tecnologias da Informação e Comunicação
UE	União Europeia
OLAP	Online Analytical Processing
DW	Data Warehouse

1. INTRODUÇÃO

1.1. ENQUADRAMENTO E IDENTIFICAÇÃO DO PROBLEMA

O rápido crescimento da densidade populacional no contexto urbano das cidades leva a uma crescente exigência sobre as infraestruturas existentes e sobre os serviços que são prestados, que devem dar resposta às necessidades e solicitações dos cidadãos.

A expansão da "Big Data" e a evolução da tecnologia associada à "Internet of Things" (IoT) tem um papel fundamental na viabilidade das iniciativas que são tomadas no contexto da temática das *Smart Cities*. A exploração da "Big Data" permite às cidades o potencial para ter acesso a muita informação, através da exploração de uma imensa quantidade de dados que chegam através de múltiplas origens e plataformas, entre as quais as plataformas associadas à IoT (Hashem et al., 2016)

A IoT consiste numa realidade em que milhões de objetos ligados à Internet comunicam entre si e geram um volume imenso de dados a cada segundo que passa. Prevê-se que estes milhões de objetos ligados entre si passem a ser os maiores produtores e consumidores de dados, no lugar do ser humano, o que significa que passa a existir uma capacidade cada vez maior de termos máquinas a processar determinados dados e a tomar decisões em nome do ser humano (Karkouch, Mousannif, Al Moatassime, & Noel, 2016)

O volume imenso de dados gerado pelas diversas origens e plataformas encaixam no conceito de "Big Data", que se caracteriza pelo volume, velocidade e a variedade dos tipos de dados que são gerados a uma taxa de crescimento cada vez maior (Hashem et al., 2016)

A decisão de implementar um sistema para tratamento de dados que chegam à organização em larga escala, de várias origens e numa velocidade cada vez maior pode transformar de uma forma fundamental a maneira como uma organização efetua a tomada de decisão. A capacidade para retirar vantagens de toda a informação disponível tem uma importância cada vez maior para o sucesso de uma organização (Janssen, van der Voort, & Wahyudi, 2017)

Com a proliferação das redes móveis, dispositivos móveis e da IoT, muitas indústrias, incluindo departamentos governamentais, empresas privadas e comunidades de pesquisa, oferecem uma maior transparência para o exterior através da divulgação de dados. O esforço resultante desse facto oferece um novo paradigma, o paradigma dos dados abertos. (Hossain, Dwivedi, & Rana, 2016)

Espera-se que os dados abertos tragam muitas vantagens, como estimular a participação cívica e a inovação, estimular a transparência e estimular o crescimento económico. Desta forma, um governo mais aberto deve ser encorajado. Vários portais e infraestruturas de dados abertos foram desenvolvidas nos últimos anos para explorar o potencial dos dados abertos, como os portais nacionais de dados abertos ou a infraestrutura europeia de dados abertos. (Zuiderwijk, Janssen, & Davis, 2014)

A crescente adoção de políticas de dados abertos por parte da administração pública, bem como os benefícios daí decorrentes, tem vindo a ser objeto de atenção das autoridades municipais na medida em que têm também um papel estruturante na construção da inteligência urbana. (de Castro Neto, Rego, Neves, & Cartaxo, 2017)

A capacidade de descobrir os dados relevantes é um pré-requisito para desbloquear o potencial dos dados abertos. Criar um portal de conjuntos de dados disponíveis é uma forma de tornar esses conjuntos de dados mais acessíveis e mais fáceis de encontrar e a extração de informação valiosa proveniente dessas diferentes fontes de dados requer a avaliação da sua qualidade. A qualidade dos dados desempenha um papel essencial no uso de portais de dados abertos e um certo nível de qualidade de dados é fundamental para a sua utilização. (Máchová & Lněnička, 2017)

Portanto, é evidente que, quando os dados que são libertados como dados abertos são de baixa qualidade, a sua reutilização será desencorajada e/ou vários utilizadores irão investir na verificação e aumento da qualidade desses dados de forma descentralizada e descoordenada: o fraco nível da documentação existente desses dados e as atividades levadas a cabo para elevar o nível de qualidade dos dados representam uma proporção significativa do esforço necessário para a reutilização dos dados abertos e isso representa um desperdício de recursos. Aumentar a qualidade dos dados abertos pode promover a sua reutilização e focar os recursos dos utilizadores em serviços de valor acrescentado. (Vetrò et al., 2016)

Neste contexto, surge o problema central que motivou esta proposta para Trabalho de Projeto para o desenvolvimento de um projeto em conjunto com a Câmara Municipal de Lisboa (CML). A CML tem neste momento vários sistemas cujos dados chegam em grandes quantidades, num grande volume, com elevada variedade e de uma forma constante. No entanto, não existe um sistema implementado que efetue uma avaliação e um tratamento desses dados, de forma a que CML possa promover a sua reutilização por parte dos cidadãos para a criação de novos produtos e serviços.

Assim sendo, existe uma grande necessidade de garantir a qualidade dos dados que são recolhidos em contextos urbanos, como por exemplo Turismo e Lazer, Transportes ou Educação.

Desta forma, de modo a poder garantir a qualidade dos dados que são colocados no Portal de Dados Abertos de Lisboa, irá ser estudada, desenvolvida e implementada uma *framework* para a avaliação e tratamento dos dados antes de serem colocados no Portal, como o objetivo principal de promover a sua reutilização por parte dos cidadãos para a criação de novos produtos e serviços.

O acesso aos dados que irão ser objeto de estudo estará garantido, uma vez que a escola tem um protocolo formal de colaboração com a CML e o Professor Miguel Neto colabora atualmente com o projeto de dados abertos da cidade.

De um modo geral, a gestão da qualidade dos dados está focada na avaliação de todos os conjuntos de dados que chegam à organização e na aplicação de ações corretivas aos dados para assegurar que se enquadram para os fins para os quais foram originalmente destinados, para que sejam úteis e apropriados. (Merino, Caballero, Rivas, Serrano, & Piattini, 2016)

1.2. OBJETIVOS DO ESTUDO

O principal impulso para o desenvolvimento deste projeto prende-se com a necessidade cada vez maior das organizações terem acesso a dados de qualidade no contexto da temática dos dados abertos. Devido a este facto, existe a necessidade de implementar metodologias de tratamento

desses dados antes da sua disponibilização, que neste caso concreto são os dados disponibilizados no Portal de Dados Abertos da cidade de Lisboa.

Assim sendo, o objetivo principal para este projeto tem como foco principal o desenvolvimento de uma *framework* para a avaliação e o tratamento da qualidade dos dados que são colocados no portal de Dados Abertos da cidade de Lisboa, de modo a promover uma maior reutilização por parte dos cidadãos para a criação de novos produtos e serviços.

Em função do objetivo principal do projeto, foram especificados e enumerados os seguintes objetivos específicos:

1. Aceder e analisar o Portal de Dados Abertos da cidade de Lisboa, de modo a construir uma matriz que contenha a identificação dos diferentes tipos de dados que o compõem;

2. Para as diferentes tipologias de dados, identificar as diferentes metodologias de controlo da qualidade dos dados a aplicar e o modo como devem ser aplicadas;

3. Testar a aplicação das metodologias e métricas identificadas para cada um dos tipos de dados identificados no Portal de Dados Abertos da cidade de Lisboa.

1.3. IMPORTÂNCIA E RELEVÂNCIA DO ESTUDO

A qualidade não depende unicamente dos dados, mas também do processo no qual os dados são recolhidos e processados. (Janssen et al., 2017)

Nos últimos anos, surgiu uma série de movimentos de dados abertos em todo o mundo, com a transparência e reutilização de dados como dois dos principais objetivos. Os Dados Abertos são dados disponibilizados gratuitamente pelos governos, organização, investigadores, entre outros, com o propósito de serem utilizados por qualquer pessoa sem restrições de direitos de autor. O crescimento do movimento de Dados Abertos tem sido muito significativo. Os Dados Abertos visam desbloquear o potencial de inovação das empresas, governos e empreendedores, mas também apresenta importantes desafios para a sua utilização efetiva. (Attard, Orlandi, Scerri, & Auer, 2015; Sadiq & Indulska, 2017)

A difusão de dados no contexto do movimento de Dados Abertos manteve um ritmo muito rápido nos últimos anos. No entanto, a experiência mostra que a divulgação de dados sem um controlo apropriado da qualidade dos mesmos pode comprometer a sua reutilização e afetar negativamente a participação cívica. (Vetrò et al., 2016)

Embora existam inúmeros sucessos de inovação baseados no paradigma de Dados Abertos, existe ainda uma incerteza quanto à qualidade dos conjuntos de dados disponibilizados, e essa incerteza representa uma ameaça ao valor que pode ser gerado a partir desses dados. (Sadiq & Indulska, 2017)

Neste caso em particular, a Câmara Municipal de Lisboa recebe de forma constante dados que são recolhidos em contextos urbanos, nas mais diversas áreas como por exemplo, Turismo e Lazer, Transportes ou Educação, entre outras.

Neste momento, a CML colabora com um conjunto de outras organizações num projeto de Dados Abertos (http://dados.cm-lisboa.pt/). Este projeto consiste num portal que disponibiliza um conjunto de dados sobre a cidade de Lisboa, produzidos pela CML e pelas entidades parceiras do programa Lisboa Aberta. O acesso aos dados é livre, pretendendo-se com isso potenciar a sua reutilização e a criação de bens e serviços que acrescentem valor aos conteúdos disponibilizados.

Atualmente, são treze as entidades que colaboram no projeto de Dados Abertos da cidade de Lisboa, que podem criar, gerir e publicar nos conjuntos de dados e esses dados podem ser catalogados em grupos, como por exemplo, Ambiente, Educação, Turismo e Lazer, entre outros, num total de dezoito grupos.

Com o desenvolvimento deste projeto, pretende-se então o desenvolvimento de uma *framework* para a avaliação e o tratamento da qualidade dos dados que são colocados no portal de Dados Abertos da cidade de Lisboa, para promover uma maior reutilização por parte dos cidadãos para a criação de novos produtos e serviços.

A implementação de um sistema de gestão de qualidade dos dados em contexto urbano pode trazer benefícios económicos e sociais à CML, na medida em que será possibilitado o acesso a uma informação que irá ajudar consideravelmente na tomada de decisão no que respeita a temas que dizem respeito a todos os cidadãos como o planeamento urbano, segurança, saúde, sistemas de transportes em termos de mobilidade e poluição, entre outros. (Rathore, Ahmad, Paul, & Rho, 2016)

1.4. METODOLOGIA

A metodologia escolhida para o desenvolvimento deste projeto é a metodologia denominada como "*Design Science Research Methodology*" (DSRM).

Segundo (Hevner & Chatterjee, 2010), o conceito de "*Design Science Research*" é um paradigma de pesquisa em que um *designer* responde a questões relevantes para problemas humanos através da criação de artefactos inovadores, contribuindo com novo conhecimento para o conhecimento científico e esses artefactos são úteis e fundamentais para a compreensão do problema em questão.

No paradigma de "*Design Science Research*", o conhecimento e a compreensão do domínio de um problema e a sua solução são alcançados na construção e aplicação de um artefacto, portanto, este processo pode ser visto como um processo de solução para problemas. (Ardakan & Mohajeri, 2009)

No caso concreto deste projeto, os artefactos são as metodologias de controlo da qualidade dos dados que são aplicadas a cada um dos tipos de dados que compõem o Portal de Dados Abertos da cidade de Lisboa.

Segundo o estudo de (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007), o desenvolvimento da metodologia DSRM requer o desenvolvimento de um processo. Deste estudo resultou a apresentação de um modelo de processo composto por seis atividades numa sequência nominal (Figura 1.1).

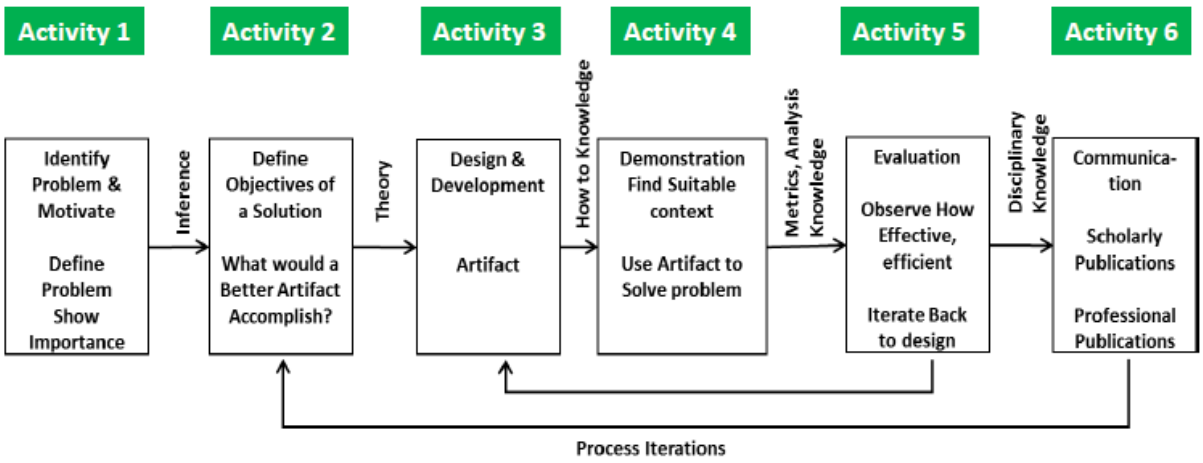

Figura 1.1 - Modelo de Processo da metodologia DSRM, retirado do estudo de (Peffers et al., 2007)

As setas que surgem na parte inferior da Figura 1.1 enfatizam a importância da iteração como parte do DSRM. Elas mostram que atividades como a Avaliação e a Comunicação resultam muitas vezes na revisão dos objetivos e desenho dos artefactos. A iteração está enraizada no processo de "*Design Science Research*" e o estudo de (Hevner & Chatterjee, 2010) ilustra esse facto com o seu *loop* de compilação e avaliação: a avaliação fornece informações de feedback sobre o artefacto projetado e uma melhor compreensão do problema que leva a uma re-iteração do processo de desenho. (Geerts, 2011)

O estudo de (Geerts, 2011) apresenta o Modelo de Processo da metodologia DSRM da Figura 1.1 explicado numa tabela. A primeira coluna da Tabela 1.1 lista as seis atividades que compõem o DSRM como uma sequência nominal. A segunda coluna descreve cada uma das atividades em detalhes: o que fazer? A terceira coluna liga a base de conhecimento às diferentes atividades: como as atividades são executadas. A base de conhecimento fornece as matérias-primas através das quais a pesquisa científica é realizada. É composta por ferramentas de conhecimento, tais como teorias, estruturas, instrumentos, construções, modelos, métodos e instâncias.

Tabela 1.1 - Tabela com a descrição das atividades que compõem o Modelo de Processo da metodologia DSRM, retirada do estudo de (Geerts, 2011)

Atividades da DSRM	**Descrição da Atividade**	**Base de Conhecimento**
Identificação do Problema e Motivação	*Qual é o problema?* Definir o problema e justificar o valor de uma solução	Compreender a relevância do problema e as suas soluções atuais com as suas fraquezas.
Definir os objetivos de uma solução	*Como resolver o problema?* Além dos objetivos gerais, como a viabilidade e o desempenho, quais são os critérios específicos que uma solução para o problema definido no ponto 1 deve atender.	Conhecimento do que é possível e o que é viável. Conhecimento de métodos, tecnologias e teorias que podem ajudar na definição dos

		objetivos.
Desenho e desenvolvimento	*Criar um artefacto que resolva o problema.* Criar modelos, métodos ou instâncias em que uma contribuição da pesquisa seja incorporada.	Aplicação de métodos, tecnologias e teorias para criar um artefacto que solucione o problema.
Demonstração	*Demonstrar o uso do artefacto.* Provar que o artefacto funciona solucionando uma ou mais instâncias do problema.	Conhecimento sobre como usar o artefacto para solucionar o problema.
Avaliação	*O artefacto funciona corretamente?* Observar e medir bem se o artefacto suporta uma solução para o problema, comparando os objetivos com resultados observados.	Conhecimento de métricas relevantes e técnicas de avaliação.
Comunicação	Comunicar o problema, a solução, a utilidade e eficácia dessa solução para os investigadores e outros públicos relevantes.	Conhecimento da cultura disciplinar.

No contexto deste Projeto, pode-se caracterizar as seis atividades que compõem o DSRM como uma sequência nominal:

1. Identificação do Problema e Motivação:

Nesta primeira etapa, irá ser justificada a importância da implementação de um sistema de avaliação da qualidade dos dados que são colocados no Portal de Dados Abertos da cidade de Lisboa. Isso será feito através da revisão da literatura que irá ajudar a ter uma melhor compreensão de como existe uma necessidade cada vez maior das organizações terem acesso a dados de qualidade no contexto da temática dos dados abertos.

A revisão de literatura também irá ajudar a verificar que existe a necessidade de implementar metodologias de avaliação da qualidade desses dados antes da sua disponibilização e quais são as metodologias aplicáveis aos diferentes tipos de dados presentes no Portal de Dados Abertos da cidade de Lisboa.

2. Definir os objetivos de uma solução:

Com base na revisão de literatura, irão ser identificados os requisitos para uma possível solução. A revisão de literatura irá permitir verificar o estado atual das *Smart Cities* com a influência cada vez maior da IoT na sua evolução, e do surgimento dos Portais de Dados Abertos. A revisão de literatura irá também permitir verificar o conceito de Dados Abertos, o seu estado atual de implementação, os seus benefícios, e a importância que a qualidade dos dados tem no sucesso dos portais de Dados Abertos e concretamente, do Portal de Dados Abertos da cidade de Lisboa.

Irá também ser efetuada uma revisão de literatura para verificar as metodologias de avaliação da qualidade dos dados existentes e a sua aplicação aos diferentes tipos de dados existentes no Portal. Assim, será possível ter as ferramentas para o desenvolvimento de uma *framework* para a avaliação e o tratamento da qualidade dos dados que são colocados no portal de Dados Abertos da cidade de Lisboa, para promover uma maior reutilização por parte dos cidadãos para a criação de novos produtos e serviços.

3. Desenho e Desenvolvimento:

Nesta fase irá ser desenhado e desenvolvido o projeto com base nos objetivos definidos no passo anterior. Este projeto tem como base o desenvolvimento de uma *framework* para a avaliação e o tratamento da qualidade dos dados que são colocados no portal de Dados Abertos da cidade de Lisboa, para promover uma maior reutilização por parte dos cidadãos para a criação de novos produtos e serviços.

Numa fase inicial irá ser efetuada uma análise do Portal de Dados Abertos da cidade de Lisboa de modo a construir uma matriz para identificar os diferentes tipos de dados existentes, e posteriormente irão ser identificadas algumas metodologias de controlo de qualidade dos dados existentes na literatura e a forma como devem ser aplicadas aos diferentes tipos de dados encontrados no Portal de Dados Abertos da cidade de Lisboa. Na última fase do desenvolvimento deste projeto será testada a aplicação das metodologias de controlo de qualidade dos dados que foram identificadas para cada tipo de dados presente no Portal.

4. Demonstração:

Nesta etapa irá ser demonstrada a aplicação das metodologias de controlo de qualidade dos dados que foram identificadas para cada tipo de dados presente no Portal de Dados Abertos da Cidade de Lisboa.

5. Avaliação:

Durante a fase de avaliação, será seguida uma abordagem exploratória. Irão ser avaliados conjuntos de dados dos vários tipos de dados identificados, com a aplicação nos dados das metodologias de controlo da qualidade dos dados identificadas. Esta avaliação exploratória irá permitir verificar o estado de alguns parâmetros de qualidade dos dados que são colocados no Portal de Dados Abertos da cidade de Lisboa, validando se este projeto suporta uma solução para o problema, comparando os objetivos com resultados observados e verificando assim se os requisitos de implementação de um sistema de gestão da qualidade dos dados que são disponibilizados no Portal de Dados Abertos da Cidade de Lisboa são cumpridos.

6. Comunicação:

As contribuições científicas das diferentes fases deste projeto poderão ser disponibilizadas publicamente no Repositório Científico de Acesso Aberto de Portugal (RCAAP) e no Repositório da Universidade Nova (RUN). Este processo é estruturado numa ordem nominalmente sequencial. Ainda assim, não existe a expectativa de que os investigadores prossigam sempre numa ordem sequencial desde a atividade 1 até à atividade 6. Na realidade, os investigadores podem começar em quase

qualquer passo e ir seguindo para os outros. Uma abordagem centrada num problema é a base da sequência nominal, começando pela atividade 1. (Peffers et al., 2007)

2. REVISÃO DA LITERATURA

No presente capítulo será apresentada a revisão de literatura que serve de base teórica para o desenvolvimento deste projeto. Esta revisão de literatura fornecerá toda a fundamentação para a elaboração de todos os modelos conceptuais para responder às necessidades do projeto.

Este capítulo inicia-se com uma breve descrição do conceito de *Smart City*, seguido de uma breve descrição do conceito de *Internet of Things* e do papel que desempenha no desenvolvimento das iniciativas que são tomadas no contexto da temática das *Smart Cities*.

Este capítulo contém também a descrição das características e objetivos do conceito de Dados Abertos, seguida da apresentação do Portal de Dados Abertos da Cidade de Lisboa, designado como Lisboa Aberta. Esta descrição é complementada também com a descrição da importância e o relevo da qualidade dos dados que são colocados no Portal, para posterior reutilização dos mesmos.

Este capítulo contém ainda uma descrição do conceito de Qualidade, com enfoque na descrição da Qualidade dos Dados, as suas características e os tipos de dados que são objeto de análise nesta disciplina, além do estudo de diversas metodologias de avaliação da qualidade dos dados.

Por fim, será também apresentado um capítulo com as características do desenho de *Dashboards* para a apresentação da informação ao utilizador final e quais são as características tipo desse utilizador final.

2.1. SMART CITIES

Existe um alto nível de acordo na literatura em como não existe ainda nenhuma definição comum sobre o que é uma *Smart City*. A crescente difusão de modelos, *standards* e definições do conceito de *Smart City* cria uma situação de ambiguidade e dificulta a estimativa sobre se as *smart cities* existentes acompanham as expectativas e os ideais reivindicados pelos promotores deste paradigma. (Grossi & Pianezzi, 2017)

No geral, a componente de IT parece ser fundamental para o conceito de uma *smart city* e os defensores deste paradigma urbano destacam os benefícios decorrentes da adoção de tecnologias, técnicas e visões que sejam científicas, objetivas, de senso comum e apolíticas. (Grossi & Pianezzi, 2017)

O rótulo "*smart city*" é um conceito ainda um pouco confuso e é utilizado numa forma que nem sempre é consistente. Não existe um enquadramento único nem uma definição única do conceito de *smart city*. (Albino, Berardi, & Dangelico, 2015)

Para se poder ter uma ideia geral sobre uma possível definição do conceito de *smart city*, é apresentada uma tabela (Tabela 2.1) que reporta algumas das diferentes definições apresentadas pelos diferentes autores na literatura.

Tabela 2.1 - Tabela adaptada com definições de Smart Cities propostas nos estudos de (Grossi & Pianezzi, 2017) e (Albino et al., 2015)

Definição	Referência
As *smart cities* representam um modelo conceptual de desenvolvimento urbano com base na utilização do capital humano, coletivo e tecnológico para a melhoria do desenvolvimento e da prosperidade nos aglomerados urbanos.	(Angelidou, 2014)
O conceito de *smart city* ocorre quando os investimentos em capital humano e social e em infraestruturas de comunicação tradicional (transportes) e moderna (TIC) fomentam um crescimento económico sustentável e um maior nível de qualidade de vida, com uma boa gestão dos recursos naturais através de uma governação participativa.	(Caragliu, Del Bo, & Nijkamp, 2011)
Uma *smart city* implica uma cidade avançada e com um alto grau tecnológico que conecta pessoas, informação e elementos da cidade recorrendo às novas tecnologias para criar uma cidade sustentável, mais verde, competitiva e com alto grau de inovação, com um consequente aumento da qualidade de vida.	(Bakici, Almirall, & Wareham, 2013)
Duas ideias principais: 1) as *smart cities* devem fazer tudo o que for relacionado com a governação e a economia utilizando novos paradigmas de pensamento e 2) as *smart cities* resumem-se a redes de sensores, dispositivos inteligentes, dados em tempo real e a integração das TIC em todos os aspetos da vida humana.	(CRETU, 2012)
As iniciativas de *Smart Cities* tentam melhorar a performance urbana utilizando dados, informação e tecnologias de informação para fornecer serviços mais eficientes aos cidadãos, para monitorizar e otimizar a infraestrutura existente, para incrementar a colaboração entre os diferentes atores económicos e para incentivar modelos de negócios inovadores em ambos os setores público e privado.	(Marsal-Llacuna, Colomer-Llinàs, & Meléndez-Frigola, 2014)

2.1.1. Smart Cities e a IoT

O surgimento da Internet das Coisas (IoT) e das Tecnologias de Informação e Comunicação (TIC) promoveu diversos conceitos e a "*Smart City*" é um desses conceitos. Tem sido um tema em foco na arena política nos últimos anos, e continua a ser, através da sua natureza específica que inclui pessoas, economia, mobilidade, ambiente, infraestruturas de TIC, estilos de vida e administração pública. (Sta, 2017)

O rápido crescimento da densidade populacional nas cidades exige que seja fornecida uma infraestrutura e serviços que vão de encontro às necessidades dos seus habitantes. Assim, tem existido um aumento no número de dispositivos, sejam eles sensores ou smartphones, levando a um potencial de negócio bastante considerável nesta nova era da Internet das Coisas (IoT), em que todos

os dispositivos estão capacitados para se interligarem e comunicarem entre eles através da Internet. (Rathore et al., 2016)

A definição de *Internet of Things* (IoT) consiste em milhões de objetos conectados e que comunicam entre si, espalhados por todo o mundo, gerando enormes quantidades de dados usando os seus sensores a cada segundo. Os dispositivos que se encontram interligados entre si irão tornar-se os maiores consumidores e produtores de dados em vez dos humanos. (Karkouch et al., 2016)

Na atualidade, estima-se que cerca de 15 bilhões de dispositivos estão conectados e está previsto que este número irá disparar para 50 bilhões em 2020, particularmente nos centros urbanos e em torno deles. As *Smart Cities* estão a abraçar as tecnologias da *Internet of Things* (IoT) por todo o mundo, de modo a agilizar as suas operações a ir de encontro às expectativas cada vez maiores dos seus cidadãos. Nos dias de hoje, os cidadãos nas cidades mais vibrantes estão já a verificar o surgimento de cada vez mais iniciativas destinadas a tornar os serviços urbanos mais inteligentes, seja para transporte, estacionamento, iluminação, tráfego e gestão de resíduos, segurança ou nas forças policiais. (Bellavista et al., 2017)

2.1.2. Smart Cities e os Dados Abertos

As *Smart Cities* visam melhorar o serviço que prestam aos seus cidadãos, tanto em termos de economia com um maior grau de eficiência, como em termos sociais, através de um maior conhecimento sobre as necessidades e desejos de todos os intervenientes. A realização desses objetivos não depende apenas dos dados fornecidos pelos governos e decisores, mas também pelos dados fornecidos pelos cidadãos, uma vez que podem ser vistos como sensores inteligentes. Neste sentido, todos os dados recolhidos são úteis para tomar decisões mais cedo e melhores, proporcionando melhores serviços aos cidadãos. (Aguilera, Peña, Belmonte, & López-de-Ipiña, 2016)

Um aspeto crucial para mudar e manter as cidades inteligentes é o uso de dados e informação, uma vez que são a base da maior parte dos serviços, ou podem até ser o próprio serviço. As novas tecnologias estão a permitir a utilização dos dados gerados por organizações públicas e a construção de serviços automatizados que respondem a questões ou problemas relacionados com a administração pública e, apesar dos dados necessários nem sempre estarem disponíveis numa forma fácil de utilizar, têm um grande potencial para criar serviços que melhorem a vida dos cidadãos e o funcionamento da sociedade. No entanto, ainda não é claro como é que os governos interagem com as partes interessadas para fornecer serviços e informações que se encaixem naquilo que as pessoas realmente querem. Existe ainda uma falta de atenção naquilo que são as necessidades e questões dos cidadãos, pelo que os dados abertos são uma forma de mitigar a separação comum entre as organizações públicas e os cidadãos. (Pereira, Macadar, Luciano, & Testa, 2017)

As multiplicidades de dados armazenados pelas administrações públicas tornaram-se uma enorme fonte de informações devido ao seu volume e diversidade. No entanto, foi apenas com o surgimento dos portais de dados abertos que esse enorme volume de dados permitiu a criação de novos modelos de negócio por partes de todas as partes envolventes numa *Smart City*. (Zotano & Bersini, 2017)

Segundo o estudo de (de Castro Neto et al., 2017), as estratégias de dados abertos permitem às cidades alcançar quatro objetivos chave:

1. Transparência: permitir que os cidadãos tenham acesso a mais informação, de modo a que possam entender, examinar e questionar as decisões que são tomadas.

2. Participação: dar acesso aos cidadãos e às suas comunidades a dados relacionados com o seu município contribui para incentivar a uma participação mais ativa e informada.

3. Melhoria do serviço e ganhos de eficiência: o fornecimento de dados abertos e a sua partilha proporciona resultados expectáveis ao nível da melhoria de serviços e de ganhos de eficiência.

4. Desenvolvimento económico: a libertação dos dados abertos permitiu às empresas e aos desenvolvedores criar novas aplicações, novos produtos e serviços, levando assim a uma promoção da atividade económica e comunitária.

2.2. DADOS ABERTOS

Ao longo dos últimos anos, os governos de todo o mundo começaram a desenvolver e a implementar iniciativas de dados abertos para permitir a divulgação de dados em formatos abertos e reutilizáveis, sem restrições ou cobrança de dinheiro pela sua utilização por parte da sociedade. Como resultado, tem surgido em todo o mundo um grande número de repositórios, catálogos e portais de dados abertos. (Máchová & Lněnička, 2017)

A proliferação e disponibilização de conjuntos de dados tornados públicos e o surgimento de mercados de dados apresentam uma oportunidade sem precedentes para governos, empresas e empresários para aproveitar o valor desses dados para conseguir ganhos económicos, sociais e científicos. (Sadiq & Indulska, 2017)

A reutilização de dados abertos promove um efeito económico positivo na inovação e no desenvolvimento de numerosas ferramentas para aumentar a eficiência, não apenas no setor privado, mas também na administração pública. (de Castro Neto et al., 2017)

Segundo (Pereira et al., 2017), os dados devem estar disponíveis para todos os que tenham um propósito para a sua utilização, e podem ser acedidos, modificados e reutilizados para qualquer finalidade. Isso significa que os dados abertos devem estar disponíveis e acessíveis, e devem poder ser reutilizados e redistribuídos para que se consiga uma participação universal, ou seja, para que toda a gente possa utilizar os dados abertos sem discriminação por campos, pessoas ou grupos.

Uma importante condição subjacente no funcionamento de uma democracia é o acesso à informação. Cidadãos informados são mais capazes de contribuir para processos democráticos, têm uma maior capacidade de compreender e aceitar as bases das decisões que os afetam e têm uma maior capacidade de se adaptarem às situações que vão ocorrendo no dia-a-dia. Nesse sentido, podemos verificar na literatura (Attard et al., 2015; Jetzek, 2016; Sieber & Johnson, 2015; Thorsby, Stowers, Wolslegel, & Tumbuan, 2017; Verhulst et al., 2016) que existem diversas definições que apontam às plataformas de dados abertos o papel de promoção dos processos democráticos e da transparência das decisões através da publicação de conjuntos de dados governamentais e

oferecendo a oportunidade de participar ativamente nos processos de decisão e resolução de problemas públicos. As plataformas de dados abertos visam também estimular a inovação, o crescimento económico e melhorar a prestação de serviços. (Ruijer, Grimmelikhuijsen, & Meijer, 2017)

Em particular, com o acesso aos dados abertos, espera-se que o público possa usar os dados do governo para tomar melhores decisões e melhorar a sua qualidade de vida, enquanto se espera que os governos possam ter acesso a um conjunto de dados com um alcance cada vez mais amplo para promover a tomada de decisão baseada em evidências. (Ubaldi, 2013)

Existe um pressuposto de que os dados abertos criam e geram mais valor do que a venda de *data sets*. Os seus benefícios podem ser agrupados em benefícios políticos e sociais, económicos e operacionais e técnicos (Tabela 2.2). (Janssen, Charalabidis, & Zuiderwijk, 2012)

Segundo (Carrara, San Chan, Fischer, & van Steenbergen, 2015), os efeitos da utilização e reutilização de dados abertos podem ser traduzidos em benefícios diretos e indiretos. Os benefícios diretos são benefícios traduzidos em valores monetários e são realizados em transações de mercado sob a forma de receitas, número de empregos envolvidos na produção de um produto ou serviço ou na poupança que podem gerar. Os benefícios indiretos da geração e reutilização de dados abertos podem ser divididos em benefícios económicos, políticos e sociais. Os benefícios económicos podem ser traduzidos em potenciais novos empregos, novos produtos e serviços, o crescimento da economia do conhecimento e maior eficiência nos serviços públicos. A nível político, os benefícios podem ser traduzidos em maior transparência e responsabilidade, participação cívica, consciencialização política e acesso à informação. Do ponto de vista social, os benefícios podem ser assumir uma forma de maior inclusão social, participação cívica, acesso à informação e suporte à tomada de decisão.

Tabela 2.2 - Tabela com alguns dos benefícios dos dados abertos, proposta no estudo de (Janssen et al., 2012)

Categoria	**Benefícios**
Políticos e Sociais	Maior Transparência, Responsabilidade Democrática, Maior participação dos cidadãos, Criação de confiança no Governo, Maior escrutínio dos dados, Igualdade no acesso aos dados, Novos serviços governamentais para os cidadãos, Melhoria dos serviços prestados aos cidadãos, Melhoria da satisfação dos cidadãos, Melhoria dos processos de formulação de políticas, Maior visibilidade para o fornecedor dos dados, Estimulação para novo conhecimento, Criação de novas visões no setor público, Novos serviços sociais (inovativos)

Económicos	Crescimento económico e estimulação da competitividade, Estimulação da inovação, Contribuição para a melhoria de processos, produtos e/ou serviços, Desenvolvimento de novos produtos e serviços, Utilização da inteligência do coletivo, adicionando assim valor à economia, Disponibilidade de informação para empresas e investidores
Operacionais e Técnicos	A capacidade de reutilizar dados / não ter que coletar os mesmos dados novamente e contrariar a duplicação desnecessária e os custos associados (também por outras instituições públicas), Otimização de processos administrativos, Melhoria de políticas públicas, Acesso a capacidade externa de resolução de problemas, Processo de tomada de decisão mais justo ao permitir a comparação, Acesso mais fácil aos dados e descoberta de dados, Criação de novos dados com base na combinação de dados, Verificações externas de qualidade de dados (validação), Sustentabilidade de dados (sem perda de dados), A capacidade de juntar e integrar dados públicos e privados

No âmbito da '*Digital Agenda for Europe*', umas das iniciativas da estratégia Europa 2020, a Comissão Europeia está focada na criação de valor através da reutilização dos dados do sector público, nomeadamente toda a informação que os organismos públicos produzem, recolhem ou pagam, incluindo, por exemplo, informação geográfica, estatística, dados meteorológicos, dados de projetos de investigação com financiamento público, etc. (de Castro Neto et al., 2017)

A União Europeia tem levado a cabo estudos para aferir a maturidade da implementação das iniciativas de dados abertos nos 28 países que a compõem. Segundo o estudo de (Carrara, Radu, & Vollers, 2017), para aferir esse grau de maturidade foram utilizados dois indicadores-chave, que se designam como '*Open Data Readiness*' e '*Portal Maturity*'. O indicador de '*Open Data Readiness*' avalia em que medida os países têm uma política de dados abertos em curso, através de normas de licenciamento, a definição de abordagens comuns e o impacto dos dados abertos. O indicador de '*Portal Maturity*' avalia a usabilidade dos portais de dados abertos, tendo em conta a disponibilidade das funcionalidades, a reutilização global dos dados e a acessibilidade dos conjuntos de dados.

Os resultados deste estudo mostram que o movimento de dados abertos ganhou uma aceitação cada vez maior a nível europeu nos últimos anos, e verifica-se que cada vez mais países europeus implementam políticas de dados abertos (Tabela 2.3).

Tabela 2.3 – Tabela com o grau de maturidade da implementação das iniciativas de dados abertos a nível da União Europeia, adaptada do estudo de (Carrara et al., 2017)

Ano	Aceitação dos Dados Abertos na União Europeia a 28
2017	73%
2016	59%
2015	44%

Detalhando o gráfico da Tabela 2.3, podemos verificar de um modo mais detalhado o crescimento do grau de maturidade através dos seus indicadores-chave através da Tabela 2.4.

Tabela 2.4 – Tabela com o grau de evolução dos dois indicadores-chave que compõem a metodologia que apura o grau de Maturidade de Dados Abertos na União Europeia, adaptada do estudo de (Carrara et al., 2017)

Ano	Prontidão dos Dados Abertos	Maturidade dos Portais de Dados Abertos
2017	72%	76%
2016	57%	66%
2015	47%	32%

Segundo (Carrara et al., 2017), o índice da maturidade global agrupa os países em diferentes clusters: *'Beginners'*, *'Followers'*, '*Fast-Trackers*' e *'Trendsetters'*. A Tabela 2.5 indica que, em 2017, o número de tendências na UE28 quase dobrou para 14 países, em comparação com apenas 8 países da UE em 2016. Os países europeus foram avaliados tanto em termos de disponibilidade de dados abertos, como na avaliação do alcance das suas políticas de dados abertos e em termos de maturidade do portal de dados abertos.

Tabela 2.5 - Tabela com a indicação e evolução da maturidade dos clusters de Dados Abertos, indicada no estudo de (Carrara et al., 2017)

Cluster	Perfil	Nº 2017	Nº 2016	Nº 2015

Beginners	O país demonstra estar numa fase inicial de maturidade em ambas as dimensões, com uma política de dados abertos em curso, assim como um portal com as funcionalidades básicas e um baixo número de conjuntos de dados. O nível de reutilização também é baixo.	1	3	7
Followers	O país já possui uma política ainda não muito profunda de dados abertos em funcionamento, bem como um portal de dados abertos com funcionalidades que vão além das funcionalidades básicas. Ainda existem limitações visíveis em termos de publicação e reutilização.	8	12	14
Fast-trackers	O país realizou progressos substanciais em termos de dados abertos, com o progresso feito numa ou em ambas as dimensões de maturidade. Ainda existem algumas barreiras em termos de disponibilização e reutilização.	8	8	-
Trendsetters	O país possui uma política avançada de dados abertos, indo além da legislação da União Europeia, bem como um portal sofisticado de dados abertos. O país possui um ecossistema de dados abertos e não demonstra limitações consideráveis, quer seja na divulgação ou na reutilização dos dados.	15	8	10

Quanto à situação em Portugal, foi aprovada recentemente uma política de dados abertos e o país está a seguir uma transição para um modelo descentralizado, onde existe um portal que irá servir como um catálogo. Além disso, vão ser selecionados setores muito específicos, para promover a qualidade ao invés da quantidade. (Carrara et al., 2017)

Em Portugal, o organismo competente por uma política nacional de dados abertos é a "Agência para a Modernização Administrativa", que promoveu o desenvolvimento de um Portal de Dados Abertos (https://dados.gov.pt/pt/), que disponibiliza vários *datasets*, provenientes de diferentes organismos ou entidades, que podem ser acedidos e descarregados por qualquer cidadão. (AMA - Agência para a Modernização Administrativa, n.d.; de Castro Neto et al., 2017)

Segundo (Ubaldi, 2013), existem quatro etapas que podem caracterizar a criação de dados abertos:

1. Geração de dados: esta fase abrange a geração de dados públicos, que normalmente é efetuada por entidades do setor público, apesar desta função ser cada vez mais compartilhada com outras fontes de dados financiadas por fundos públicos (por exemplo, estatísticas sociais, dados de tráfego aéreo).

2. Recolha, Agregação e Processamento de dados: os dados necessitam de ser recolhidos e reunidos para permitir o acesso, partilha e reutilização dos mesmos. A maioria dos dados governamentais necessitam também de serem agregados, ligados e/ou manipulados assim que são acedidos, de modo a que lhes seja dado um valor que permita suportar a tomada de

decisão. Muitos utilizadores não são capazes de entender e utilizar os dados que chegam em bruto, ou seja, sem qualquer tipo de manipulação.

3. Distribuição e Entrega de dados: os dados necessitam de ser distribuídos aos potenciais utilizadores para permitir o acesso e a sua reutilização

4. Utilização final de dados: os dados abertos do governo necessitam de ser reutilizados por um conjunto de utilizadores diferentes, de modo a sustentar a criação de valor público.

Efetuar a publicação de dados abertos a partir de uma fonte de dados é um processo que se sobrepõe a processos de armazenamento de dados em que os dados são extraídos, transformados e carregados (processo ETL). No caso da publicação de dados abertos, a fase de carregamento é substituída pela fase de publicação (processo ETP - *Extract*, *Transform*, *Publish*). O processo ETP (Figura 2.1) é a especificação técnica de como os dados fluem através da organização, como são transformados em conjuntos de dados publicáveis e como, eventualmente, são tornados públicos.

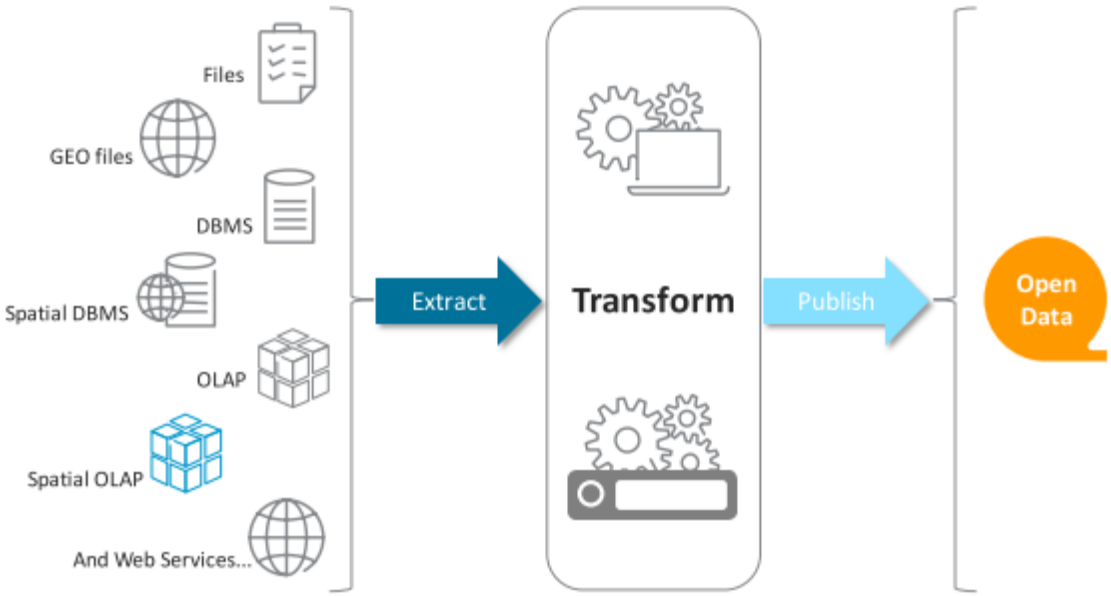

Figura 2.1 – Visualização do Processo de *Extract*, *Transform*, *Publish* (ETP), retirada do estudo de (Carrara, Oudkerk, Steenbergen, & Tinholt, 2016)

2.2.1. Portal de Dados Abertos da Cidade de Lisboa

O movimento de dados abertos tem como objetivo abrir a informação do sector público com o propósito de maximizar a sua reutilização. Uma implementação típica baseia-se na coleta e na publicação dos conjuntos de dados em portais centrais ou catálogos de dados, de modo a fornecer um "*one-stop-shop*" para os consumidores de dados. Enquanto o funcionamento comum de um catálogo de dados baseia-se no registo de fontes de dados através do fornecimento de links, um

portal funciona como um ponto de entrada único que contém os dados, onde os utilizadores podem efetuar pesquisas, podem aceder aos dados publicados e explorar o portal. (Attard et al., 2015)

Seguindo este princípio, existe uma iniciativa de dados abertos da Câmara Municipal de Lisboa (CML) de modo a disponibilizar dados para a população. A CML contém um grande conjunto de dados que, quando compartilhados, podem criar valor e é nesse sentido que tem tornado público esses conjuntos de dados, através de um projeto de Dados Abertos (http://lisboaaberta.cm-lisboa.pt/index.php/pt/). Este projeto consiste num portal que disponibiliza um conjunto de dados sobre a cidade de Lisboa, produzidos pela CML e pelas entidades parceiras do programa Lisboa Aberta, uma vez que a CML colabora com um conjunto de outras organizações. O acesso aos dados é livre, pretendendo-se com isso potenciar a sua reutilização e a criação de bens e serviços que acrescentem valor aos conteúdos disponibilizados.

Atualmente, são treze as entidades que colaboram no projeto de Dados Abertos da cidade de Lisboa, que podem criar, gerir e publicar nos conjuntos de dados, num total de 336 (Tabela 2.6).

Tabela 2.6 - Tabela com indicação das Entidades que colaboram no Projeto de Dados Abertos de Lisboa e o nº de Conjuntos de Dados no Portal (Câmara Municipal de Lisboa, n.d.)

Entidade	**Nº de Conjuntos de Dados no Portal**
Associação de Turismo de Lisboa	4
Câmara Municipal de Lisboa	247
EDP	2
EGEAC	1
Emel	5
Gebalis	2
Instituto Hidrográfico	3
INE - Instituto Nacional de Estatística	50
IPMA – Instituto Português do Mar e Atmosfera	5
Lisboa E-Nova	2
SRU Ocidental	1
Transporlis	9
Turismo de Portugal	5

A publicação de dados no Portal de Dados Abertos de Lisboa utiliza um sistema de catálogo '*open source*', denominado CKAN (*Comprehensive Knowledge Archive Network*), que se tornou uma das principais plataformas em portais de dados abertos (https://ckan.org/). A plataforma CKAN é um poderoso sistema de gestão de dados que torna os dados acessíveis, fornecendo ferramentas para efetuar a publicação, partilha, pesquisa e utilização dos dados. Como a plataforma CKAN é '*open source*', é continuamente melhorada pela comunidade ao longo do tempo e está disponível gratuitamente.

A gestão do portal de dados abertos de Lisboa é assegurada pela Divisão de Gestão de Informação Georreferenciada da Câmara Municipal de Lisboa e a utilização dos dados disponibilizados é regida por licenças recomendadas pela *Open Knowledge International* (https://okfn.org/). Segundo o estudo de (de Castro Neto et al., 2017), a OKFN é uma rede mundial sem fins lucrativos e é uma das organizações que trabalham para promover os dados abertos, promovendo-os, usando a tecnologia, metodologia e boas-práticas para libertar dados e permitir que as pessoas trabalhem com eles para criar e compartilhar conhecimento.

Na Figura 2.2 é possível verificar o aspeto geral do Portal de Dados Abertos da Cidade de Lisboa.

Figura 2.2 - Página principal do Portal de Dados Abertos da Cidade de Lisboa

Como é possível verificar, os 336 conjuntos de dados podem ser subdivididos por 18 grupos:

1. Administração Pública e Justiça
2. Ambiente
3. Cultura e Património
4. Educação
5. Segurança e Socorro

6. Desporto
7. Gestão Urbana
8. Economia e Inovação
9. Energia e Comunicações
10. Planeamento Urbano
11. Habitação e Desenvolvimento social
12. Informação Base e Cartografia
13. População
14. Outros Equipamentos e Serviços
15. Saúde
16. Transparência
17. Turismo e Lazer
18. Mobilidade

De acordo com o estudo de (Lourenço, 2015), existem uma série de características chave que um portal de dados abertos deve respeitar de modo a que o portal seja uma plataforma que promova a transparência e a responsabilidade:

1. Qualidade

   Os cidadãos esperam que os dados divulgados pelas entidades oficiais tenham qualidade no sentido em que são dados oficiais e, portanto, devem ser dados precisos e confiáveis.

2. Integridade dos dados

   Ao publicar um determinado conjunto de dados, as entidades públicas devem considerar o nível de detalhe em que esses dados são divulgados. Dados demasiado detalhados podem comprometer sua compreensibilidade e "esconder" os dados relevantes, mas a falta de detalhes pode criar conjuntos de dados incompletos.

3. Acesso e Visibilidade

   Devem existir poucas ou nenhumas barreiras ao acesso dos cidadãos aos dados relevantes

4. Usabilidade e Compreensão

   A usabilidade dos portais de dados abertos refere-se à facilidade com que os utilizadores conseguem aceder às informações e navegar no portal. A compreensão dos portais de dados abertos visa garantir que os utilizadores conseguem entender e interpretar adequadamente

a informação a que acedem no portal de dados abertos. Os dados não devem conter significados ambíguos ou jargão.

5. Temporalidade dos dados

   Os dados colocados no portal de dados abertos devem ser "atuais" e devem ser atualizados ao longo do tempo, de modo a promover a sua usabilidade por parte dos utilizadores.

6. Valor e utilidade

   As entidades públicas não devem tentar publicar todos os dados que possuem. Devem começar por identificar os dados mais relevantes com mais alto valor e maior impacto que mais beneficiam os utilizadores.

7. Granularidade

   Os dados devem ser publicados no portal de dados abertos como chegaram da fonte, ou seja, com um grau de granularidade mínimo, sem agregações ou alterações. Os dados publicados não devem dar aos utilizadores uma visão pré-formatada ou tendenciosa.

2.2.2. Qualidade dos Dados Abertos

A difusão das iniciativas de dados abertos nos últimos anos manteve um ritmo muito elevado. No entanto, existem evidências de que a divulgação de dados sem um controlo apropriado da qualidade pode comprometer a reutilização dos mesmos e afetar negativamente a participação cívica. (Vetrò et al., 2016)

Embora existam inúmeros sucessos de inovação baseados no paradigma dos dados abertos, existe alguma incerteza sobre a qualidade dos dados nos conjuntos de dados disponibilizados. Essa incerteza é uma ameaça ao valor que pode ser gerado a partir desses dados. (Sadiq & Indulska, 2017)

A qualidade dos dados pode ser definida como um grau em que um conjunto de características dos dados cumpre com determinados requisitos. Essas características podem ser: integridade dos dados, validade, precisão, consistência, disponibilidade e atualidade dos dados. Os requisitos são definidos conforme a necessidade ou expectativa que é declarada, geralmente implícita ou obrigatória. O standard ISO 9001:2015 considera a qualidade dos dados como um conceito relativo, largamente dependente de requisitos específicos que resultam da utilização dos dados. Isto significa que os mesmos dados podem ser de boa qualidade para um tipo de utilização e completamente inúteis para outro. (Bicevskis, Bicevska, & Karnitis, 2017)

O estudo de (Batini, Cappiello, Francalanci, & Maurino, 2009) apresenta uma análise das abordagens existentes para a avaliação da qualidade dos dados e a identificação de requisitos, indicando que tais abordagens incluem tipicamente três aspetos principais: análise dos dados e dos processos, análise dos requisitos da qualidade dos dados e análise da qualidade dos dados.

A análise dos dados e dos processos inclui o estudo dos *schemas* dos dados, a realização de reuniões com os utilizadores para que se possa alcançar uma compreensão completa dos dados, as suas regras e restrições e os processos de criação e consumo desses dados. A análise de requisitos da

qualidade dos dados inclui geralmente sondagens aos utilizadores e administradores dos dados para que se possam identificar problemas na qualidade dos dados, com o objetivo específico de identificar conjuntos de dados críticos, definir métricas para medir a qualidade dos dados e definir metas. A análise da qualidade dos dados refere-se às atividades relacionadas com a exploração e avaliação dos conjuntos de dados em relação às métricas da qualidade de dados definidas anteriormente. (Batini et al., 2009)

À medida que a quantidade e a variedade das fontes de dados vão aumentando, é importante criar bons metadados (descrições, cobertura geográfica, limitações, etc.), a fim de permitir que as partes interessadas, que não têm um grande conhecimento sobre o assunto, facilmente pesquisem e consumam os dados. Em certo sentido, a noção de que todos os dados divulgados devem ter qualidade é evidente, mas esse conceito não é fácil de identificar no contexto dos dados abertos, e o requisito para a qualidade dos dados pode ser considerado como abrangendo diversas características. (Máchová & Lněnička, 2017)

Os cidadãos esperam que os dados divulgados pelas entidades oficiais tenham qualidade no sentido em que são dados oficiais e, portanto, devem ser precisos e confiáveis. Por vezes não é claro qual a organização em particular que é responsável por um portal, e mesmo quando isso é conhecido, os cidadãos podem não estar cientes da sua credibilidade. Além disso, quando os dados apresentados foram recolhidos e processados por entidades externas, os portais podem não ser diretamente responsáveis pela sua qualidade. Por outro lado, os cidadãos comuns não possuem os conhecimentos necessários para avaliar, por si mesmos, a qualidade dos dados divulgados, que podem variar de acordo com cada conjunto de dados específico. (Lourenço, 2015)

A qualidade da informação pode variar e ser muito baixa. Os cidadãos esperam que o governo seja responsável pela qualidade dos dados. A abertura de dados que não possuem uma qualidade de informação adequada pode resultar em discussões, confusões, menos transparência, e em menos confiança no próprio governo. Este último ponto pode ser explicado pelo facto da baixa qualidade da informação resultar em recursos que são desperdiçados e em resultados confusos ou incorretos. O ditado "*garbage in, garbage out*" é realmente verdadeiro no que toca aos dados abertos. (Janssen et al., 2012)

Os decisores políticos preferem simplesmente disponibilizar os dados, sem se preocuparem com a proveniência ou o enriquecimento dos mesmos. Este mito indica que os dados podem ser disponibilizados sem quaisquer atividades adicionais. Os dados de origem geralmente não podem ser utilizados de imediato e a avaliação da qualidade e o processamento dos dados são fundamentais para a sua utilização. (Janssen et al., 2012)

Um dos maiores riscos relacionados com a utilização dos dados abertos prende-se com a falta de consciência da qualidade dos dados. Os consumidores de dados abertos normalmente não são os produtores e, portanto, não existe uma estratégia bem definida para a limpeza dos dados, que geralmente resulta numa má limpeza e transformação dos mesmos. Os consumidores de dados abertos podem, portanto, investir esforços significativos para gerar resultados valiosos a partir dos dados, apenas para perceber que os resultados são inadequados, ou não perceberem que os dados são de má qualidade e confiar em resultados incorretos. (Sadiq & Indulska, 2017)

As medidas sobre a qualidade dos dados podem ser aplicadas no domínio dos dados abertos. O sucesso dos dados abertos depende fortemente da qualidade dos conjuntos de dados disponibilizados, uma vez que existe uma grande variedade na qualidade dos conjuntos de dados libertados e os utilizadores também podem ter uma preocupação com a qualidade dos dados abertos. Experiências recentes mostram também que a qualidade dos registos do catálogo de dados pode afetar a capacidade dos utilizadores de localizar os dados do seu interesse. (Máchová & Lněnička, 2017)

A literatura fornece uma ampla gama de técnicas para avaliar e melhorar a qualidade dos dados, tais como as ligações entre registos, regras de negócio ou medidas de similaridade. Ao longo do tempo, essas técnicas evoluíram para lidar com a crescente complexidade da qualidade de dados em sistemas de informação em rede. Devido à diversidade e complexidade dessas técnicas, a pesquisa concentrou-se na definição de metodologias que ajudam a selecionar, personalizar e aplicar técnicas de avaliação e melhoria da qualidade dos dados. (Batini et al., 2009)

2.2.3. Gestão da Qualidade dos Dados

De acordo com a norma ISO 9000:2005, o conceito de Qualidade define-se pelo grau com que um conjunto de características inerentes cumpre os requisitos. Segundo a mesma norma, o conceito de Requisito define-se pela necessidade ou expectativa que é declarada, geralmente implícita ou obrigatória. Assim sendo, podemos verificar que a Qualidade dos Dados se define por dados que vão de encontro aos requisitos dos utilizadores.

Segundo o *standard* (ISO 9001, 2008), a adoção de um sistema de gestão da qualidade deverá ser uma decisão estratégica da organização. A conceção e a implementação do sistema de gestão da qualidade de uma organização é influenciada:

a) pelo seu ambiente organizacional, por mudanças nesse ambiente e por riscos associados a esse ambiente;

b) por necessidades variáveis;

c) por objetivos particulares;

d) pelos produtos que proporciona;

e) pelos processos que utiliza;

f) pelas suas dimensões e estrutura organizacional.

Esta norma fomenta a adoção de uma abordagem por processos quando se desenvolve, implementa e melhora a eficácia de um sistema de gestão da qualidade, para aumentar a satisfação do cliente ao ir ao encontro dos seus requisitos.

A adoção de uma abordagem por processos tem como base uma atividade ou conjunto de atividades utilizando recursos, gerida de forma a permitir a transformação de entradas em saídas e em que a saída de um processo constitui diretamente a entrada do seguinte. Uma vantagem da abordagem

por processos é o controlo passo-a-passo que proporciona sobre a interligação dos processos individuais dentro do sistema de processos, bem como sobre a sua combinação e interação.

Quando utilizada dentro de um sistema de gestão da qualidade, tal abordagem enfatiza a importância:

a) de entender e ir ao encontro dos requisitos;

b) da necessidade de considerar processos em termos de valor acrescentado;

c) de obter resultados do desempenho e da eficácia do processo;

d) da melhoria contínua dos processos baseada na medição dos objetivos.

Neste modelo de um sistema de gestão da qualidade baseado em processos, os clientes têm um papel significativo na definição de requisitos como entradas e a monitorização da satisfação do cliente requer a avaliação da informação relativa à perceção, por parte deste, quanto à organização ter ido ao encontro dos seus requisitos.

Segundo esta norma ISO 9001:2008, a gestão de topo deve assegurar que a política de Qualidade:

a) é apropriada ao propósito da organização;

b) inclui um comprometimento de cumprir os requisitos e de melhorar continuamente a eficácia do sistema de gestão da qualidade;

c) proporciona um enquadramento para o estabelecimento e a revisão dos objetivos da qualidade;

d) é comunicada e entendida dentro da organização;

e) é revista para se manter apropriada.

A gestão de topo deve assegurar que os objetivos da qualidade são mensuráveis e consistentes com a política da Qualidade.

A organização deve planear e desenvolver os processos necessários para a realização do produto. No planeamento da realização do produto, a organização deve determinar, conforme apropriado, o seguinte:

a) objetivos da qualidade e requisitos para o produto;

b) a necessidade de estabelecer processos e documentos, e de proporcionar os recursos específicos para o produto;

c) as atividades requeridas de verificação, validação, monitorização, medição, inspeção e ensaio específicas do produto e os critérios de aceitação do produto;

d) os registos necessários para proporcionar a evidência de que os processos de realização e o produto resultante vão de encontro aos requisitos

2.2.4. Metodologias para avaliação e melhoria da qualidade dos dados

As consequências da má qualidade dos dados são verificadas muitas vezes na vida quotidiana das empresas, mas, na maioria das vezes não são efetuadas as conexões necessárias para apurar as causas. A qualidade dos dados tem sérias consequências para a eficiência e eficácia das empresas. (Batini, Barone, Mastrella, Maurino, & Ruffini, 2007)

Para as empresas, garantir a qualidade dos dados é uma tarefa múltipla que exige que os dados sejam atualizados, completos, consistentes, válidos e acessíveis para melhorar as interações com o cliente. Dado este desafio global e recorrente e num contexto de recursos escassos, os erros nos dados precisam de ser abordados por metodologias orientadas por um sentido de economia. Ou seja, para uma melhoria na qualidade dos dados devem ser escolhidas as metodologias de acordo com a sua eficácia e, em seguida, aplicadas numa sequência eficiente. (Kleindienst, 2017)

Existem várias metodologias que foram propostas na literatura, com conjuntos de métricas, características e dimensões para avaliar a qualidade de conjuntos de dados. Uma metodologia de avaliação da qualidade dos dados é definida como o processo de avaliação da adequação dos conjuntos de dados às informações que os utilizadores necessitam num caso específico. O processo envolve a medição das dimensões de qualidade relevantes para o utilizador e, posteriormente comparando os resultados da avaliação com os requisitos de qualidade do utilizador. (Rula & Zaveri, 2014; Vetrò et al., 2016)

De acordo com o estudo de (Batini et al., 2009), a sequência de atividades de uma metodologia de qualidade de dados é composta por três fases:

1. Reconstrução do Estado da situação: que visa a obtenção de informações contextuais sobre processos e serviços organizacionais, obtenção de dados e procedimentos de gestão relacionados, e questões relacionadas com a qualidade e custos correspondentes.

2. Avaliação / medição: que mede a qualidade da coleta de dados ao longo de algumas dimensões de qualidades relevantes; no termo 'medição' é medido o valor de um conjunto de dimensões de qualidade de dados. O termo 'avaliação' é utilizado quando essas medidas são comparadas com os valores de referência, de modo a permitir um diagnóstico de qualidade.

3. Melhoria: diz respeito à seleção das etapas, estratégias e técnicas para alcançar os novos objetivos de qualidade de dados

Após uma pesquisa na literatura, na Tabela 2.7 pode-se verificar um conjunto de metodologias de avaliação e melhoria da qualidade dos dados.

Tabela 2.7 – Lista de Metodologias consideradas neste artigo

Acrónimo da Metodologia	Designação	Referência
HDQM	*Heterogenous Data Quality Methodology*	(Batini et al., 2009; Carlo, Daniele, Federico, & Simone, 2011)

TDQM	*Total Data Quality Methodology*	(Batini et al., 2009; Wang, 1998)
TIQM	*Total Information Quality Management*	(Batini et al., 2009; English, 1999)
AIMQ	*A methodology for information quality assessment*	(Batini et al., 2009; Lee, Strong, Kahn, & Wang, 2002)
CIHI	*Canadian Institute for Health Information methodology*	(Batini et al., 2009; Long & Seko, 2002)
ISTAT	*ISTAT methodology*	(Batini et al., 2009; Falorsi, P.D.; Pallara, S.; Pavone, A.; Alessandroni, A.; Massella, E.; and Scannapieco, 2003)
COLDQ	*Loshin Methodology (Cost-effect Of Low Data Quality)*	(Batini et al., 2009; Loshin, 2001)
DaQuinCIS	*Data Quality in Cooperative Information Systems*	(Batini et al., 2009; Scannapieco, Virgillito, Marchetti, Mecella, & Baldoni, 2004)
CDQ	*CDQ - Comprehensive methodology for Data Quality management*	(Batini, Cabitza, Cappiello, & Francalanci, 2006; Batini et al., 2009)
DWQ	*DWQ - The Data Warehouse Quality Methodology*	(Batini et al., 2009; Jeusfeld, Quix, & Jarke, 1998)
DQA	*Data Quality Assessment*	(Batini et al., 2009; Pipino, Lee, Wang, Lowell Yang Lee, & Yang, 2002)
IQM	*Information Quality Measurement*	(Batini et al., 2009; Eppler & Helfert, 2004)
AMEQ	*Activity-based Measuring and Evaluating of product information Quality (AMEQ) methodology*	(Batini et al., 2009; Su & Jin, 2004)
QAFD	*Methodology for the Quality Assessment of Financial Data*	(Batini et al., 2009; De Amicis, Barone, & Batini, 2006)
SPDQM	*Square-Aligned Portal Data*	(Moraga, Moraga, Calero, &

	Quality Model	Caro, 2009; Vetrò et al., 2016)
PDQM	*Portal Data Quality Model*	(Calero, Caro, & Piattini, 2008)
ORME-DQ	*ORME-DQ, a methodology and a framework for data quality assessment*	(Batini et al., 2007)
MAMD 2.0	*MAMD – Modelo Alarcos de Mejora de Datos*	(Carretero, Gualo, Caballero, & Piattini, 2017)
WIQA	*WIQA - Information Quality Assessment Framework*	(Bizer & Cyganiak, 2009)

A qualidade dos dados é normalmente definida como uma construção multidimensional com a definição popular "aptidão para o uso". A qualidade dos dados pode depender de vários fatores (dimensões ou características), tais como *Accuracy*, *Timeliness*, *Completeness*, *Relevancy*, *Objectivity*, *Believability*, *Understandability*, *Consistency*, *Conciseness*, *Availability*, *Verifiability*. (Zaveri et al., 2012)

A avaliação da qualidade dos dados envolve a medição de dimensões e critérios de qualidade relevantes para o consumidor. As dimensões podem ser consideradas como as características de um conjunto de dados. (Zaveri et al., 2012)

Na Tabela 7.1 da secção de Apêndices, pode ser consultada a descrição de todas as dimensões/características, encontradas na literatura que foi estudada.

Na Tabela 2.8 podemos verificar um mapeamento entre as metodologias presentes neste documento e as suas dimensões/características.

Tabela 2.8 – Tabela com mapeamento entre as metodologias presentes neste documento e as suas dimensões/características

Acrónimo da Metodologia	**Dimensões/Características**	**Referência**
HDQM	Accuracy, Currency	(Batini et al., 2009; Carlo et al., 2011)
TDQM	Accessibility, Appropriateness, Believability, Completeness, Concise/Consistent representation, Ease of manipulation, Value added, Free of error, Interpretability, Objectivity, Relevance, Reputation, Security, Timeliness, Understandability	(Batini et al., 2009; Wang, 1998)
TIQM	**Dimensões inerentes**: Definition conformance (consistency), Completeness, Business rules conformance, Accuracy (to	(Batini et al., 2009;

	surrogate source), Accuracy (to reality), Precision, Nonduplication, Equivalence of redundant data, Concurrency of redundant data **Dimensões Pragmáticas**: Accessibility, Timeliness, Contextual Clarity, Derivation Integrity, Usability, Rightness (fact completeness), Cost.	English, 1999)
AIMQ	Accessibility, Appropriateness, Believability, Completeness, Concise/Consistent representation, Ease of operation, Freedom from errors, Interpretability, Objectivity, Relevancy, Reputation, Security, Timeliness, Understandability	(Batini et al., 2009; Lee et al., 2002)
CIHI	Accuracy, Timeliness Comparability, Usability, Relevance, Over-coverage, Under-coverage, Simple/correlated response variance, Reliability, Collection and Capture, Unit/Item non response, Edit and imputation, Processing, Estimation, Timeliness, Comprehensiveness, Integration, Standardization, Equivalence, Linkage ability, Product/Historical comparability, Accessibility, Documentation, Interpretability, Adaptability, Value	(Batini et al., 2009; Long & Seko, 2002)
ISTAT	Accuracy, Completeness, Consistency	(Batini et al., 2009; Falorsi, P.D.; Pallara, S.; Pavone, A.; Alessandroni, A.; Massella, E.; and Scannapieco, 2003)
COLDQ	**Schema**: Clarity of definition, Comprehensiveness, Flexibility, Robustness, Essentialness, Attribute granularity, Precision of domains, Homogeneity, Identifiability, Obtainability, Relevance, Simplicity/Complexity, Semantic consistency, Syntactic consistency. **Dados**: Accuracy, Null Values, Completeness, Consistency, Currency, Timeliness, Agreement of Usage, Stewardship, Ubiquity **Apresentação**: Appropriateness, Correct Interpretation, Flexibility, Format precision, Portability, Consistency, Use of storage, **Política de informação**: Accessibility, Metadata, Privacy, Security, Redundancy, Cost.	(Batini et al., 2009; Loshin, 2001)
DaQuinCIS	Accuracy, Completeness, Consistency, Currency,	(Batini et al., 2009; Scannapieco et al.,

	Trustworthiness	2004)
CDQ	***Schema***: Correctness with respect to the model, Correctness with respect to Requirements, Completeness, Pertinence, Readability, Normalization **Dados**: Syntactic/Semantic Accuracy, Semantic Accuracy, Completeness, Consistency, Currency, Timeliness, Volatility, Completability, Reputation, Accessibility, Cost.	(Batini et al., 2006, 2009)
DWQ	Correctness, Completeness, Minimality, Traceability, Interpretability, Metadata Evolution, Accessibility (System, Transactional, Security), Usefulness (Interpretability), Timeliness (Currency, Volatility), Responsiveness, Completeness, Credibility, Accuracy, Consistency, Interpretability	(Batini et al., 2009; Jeusfeld et al., 1998)
DQA	Accessibility, Appropriate amount of data, Believability, Completeness, Freedom from errors, Consistency, Concise Representation, Relevance, Ease of manipulation, Interpretability, Objectivity, Reputation, Security, Timeliness, Understandability, Value added	(Batini et al., 2009; Pipino et al., 2002)
IQM	Accessibility, Consistency, Timeliness, Conciseness,Maintainability, Currency, Applicability, Convenience, Speed, Comprehensiveness, Clarity, Accuracy, Traceability, Security, Correctness, Interactivity	(Batini et al., 2009; Eppler & Helfert, 2004)
AMEQ	Consistent representation, Interpretability, Case of understanding, Concise representation, Timeliness, Completeness Value added, Relevance, Appropriateness, Meaningfulness, Lack of confusion, Arrangement, Readable, Reasonability, Precision, Reliability, Freedom from bias, Data Deficiency, Design Deficiency, Operation Deficiencies, Accuracy, Cost, Objectivity, Believability, Reputation, Accessibility, Correctness, Unambiguity, Consistency	(Batini et al., 2009; Su & Jin, 2004)
QAFD	Syntactic/Semantic accuracy, Internal/External consistency, Completeness, Currency, Uniqueness	(Batini et al., 2009; De Amicis et al., 2006)
SPDQM	Accessibility, Accuracy, Amount of data, Applicability, Attractiveness, Availability, Completeness, Compliance, Concise Representation, Confidentiality, Consistency, Consistent Representation, Credibility, Currentness, Customer Support, Documentation, Ease of operation, Effectiveness, Efficiency, Expiration, Flexibility, Interactive, Interpretability,	(Moraga et al., 2009; Vetrò et al., 2016)

	Novelty, Objectivity, Organization, Portability, Precision, Readability, Recoverability, Relevancy, Reliability, Reputation, Scope, Specialization, Timeliness, Traceability, Understandability, Usefulness, Validity, Value added, Verifiability	
PDQM	Attractiveness, Accessibility, Accuracy, Amount of data, Applicability, Availability, Believability, Completeness, Concise representation, Consistent representation, Currency, Documentation, Duplicates, Ease of operation, Expiration, Flexibility, Interactivity, Interpretability, Novelty, Objectivity, Organization, Relevancy, Customer support, Reliability, Reputation, Response time, Security, Specialization, Timeliness, Traceability, Understandability, Validity, Value added	(Calero et al., 2008)
ORME-DQ	Completeness, Currency, Consistency	(Batini et al., 2007)
WIQA	Accuracy, Timeliness, Completeness, Relevance, Objectivity, Believability, Understandability, Consistency, Conciseness, Availability, Verifiability	(Bizer & Cyganiak, 2009)

Uma das metodologias indicadas na Tabela 2.8, a metodologia SPDQM foi construída com base na metodologia *Portal Data Quality Model* (PDQM), do estudo de (Calero et al., 2008), e o padrão ISO/IEC 25012 (ISO/IEC 25012, 2008), o padrão de qualidade de dados que faz parte da família SQuaRE (*Software Quality product Requirements and Evaluation*). (Moraga et al., 2009)

Segundo os estudos de (Calero et al., 2008; Moraga et al., 2009; Vetrò et al., 2016), a metodologia SPDQM contém um conjunto de 42 características (30 características da metodologia PDQM, 7 características do SQUARE, e as 5 características restantes foram adicionadas após uma revisão sistemática da literatura), que são organizadas em dois pontos de vista e quatro categorias:

- Inerente
    - Intrínseco: que denota que os dados têm qualidade por si mesmos
- Dependente do Sistema
    - Operacional: que enfatiza a importância do papel dos sistemas, ou seja, o sistema deve estar acessível, mas seguro
    - Contextual: que destaca o requisito que indica que a qualidade dos dados deve ser considerada no contexto da tarefa que temos em mãos

- Representativo: que denota que o sistema deve apresentar dados de uma forma a que sejam interpretáveis, fáceis de entender e representados de uma forma concisa e consistente

Os problemas mais comuns de qualidade dos dados que são reportados correspondem a propriedades intrínsecas dos dados ou propriedades que dependem do contexto em que são utilizados. (Bovee, Srivastava, & Mak, 2003; Vetrò et al., 2016)

Segundo o estudo de (Vetrò et al., 2016), uma vez que o conceito de dados abertos abrange domínios heterogéneos e estão sujeitos a uma utilização bastante diversificada por parte dos seus consumidores, é preferível selecionar as dimensões que abordam os aspetos intrínsecos da qualidade dos dados. Deste ponto de vista, a metodologia SPDQM contém o conjunto mais completo de características (12) quando comparado com outros modelos. Além disso, a metodologia SPDQM apresenta um conjunto de características básicas compartilhadas por quase todos os modelos (*Accuracy*, *Completeness*, *Timeliness*), e fornece características como *Traceability*, *Compliance* e *Understandability*, que são características menos consideradas por outras metodologias, mas que são importantes num contexto de dados abertos.

2.2.5. Dimensões e Métricas

Em todas as metodologias, a definição das dimensões e métricas para avaliar os dados é uma atividade crítica. Em geral, podem ser associadas várias métricas a cada dimensão de qualidade. Em alguns casos, a métrica é única e a definição teórica de uma dimensão coincide com a definição operacional da métrica correspondente. (Batini et al., 2009)

Uma métrica, medida ou indicador de avaliação de qualidade de dados é um procedimento para medir uma dimensão de qualidade de dados. Essas métricas são heurísticas que são projetadas para atender a uma situação de avaliação específica. Uma vez que as dimensões são conceitos bastantes abstratos, as métricas de avaliação dependem de indicadores de qualidade que permitem a avaliação da qualidade de uma fonte de dados. A pontuação da avaliação é calculada a partir desses indicadores de qualidade, através do recurso a funções de cálculo. (Zaveri et al., 2012)

O objetivo final das funções métricas é poder determinar se os dados são adequados para o propósito, de acordo com o contexto da sua utilização. Num serviço de apuramento da qualidade dos dados, cada função métrica tem a capacidade de quantificar numericamente a qualidade dos dados apurada na sua dimensão específica, dentro do contexto da sua utilização em termos comerciais. Algumas dimensões de qualidade de dados podem ser muito subjetivas e podem ser difíceis de medir. (Yeganeh, Sadiq, & Sharaf, 2014)

De acordo com o estudo de (Heinrich, Kaiser, & Klier, 2011), para suportar uma gestão da qualidade de dados orientada com uma vertente económica, são necessárias métricas para quantificar a qualidade dos dados. O apuramento dessas métricas permite a resposta a perguntas como as seguintes: Qual a medida que melhora mais a qualidade dos dados? Qual delas tem a melhor relação custo-benefício?

A Figura 2.3 ilustra o circuito fechado de uma gestão de qualidade de dados orientada com uma vertente económica. Este circuito pode ser influenciado através de medidas de qualidade de dados (por exemplo, medidas de limpeza de dados, etc.). Tomar medidas melhora o nível atual da qualidade de dados (quantificado através da utilização de métricas) e este facto leva a um benefício económico correspondente. Do ponto de vista económico, apenas devem ser tomadas medidas que sejam eficientes em relação aos custos e benefícios. Por exemplo, se existirem duas medidas com o mesmo benefício económico, é racional escolher aquela com custos mais baixos. (Heinrich et al., 2011)

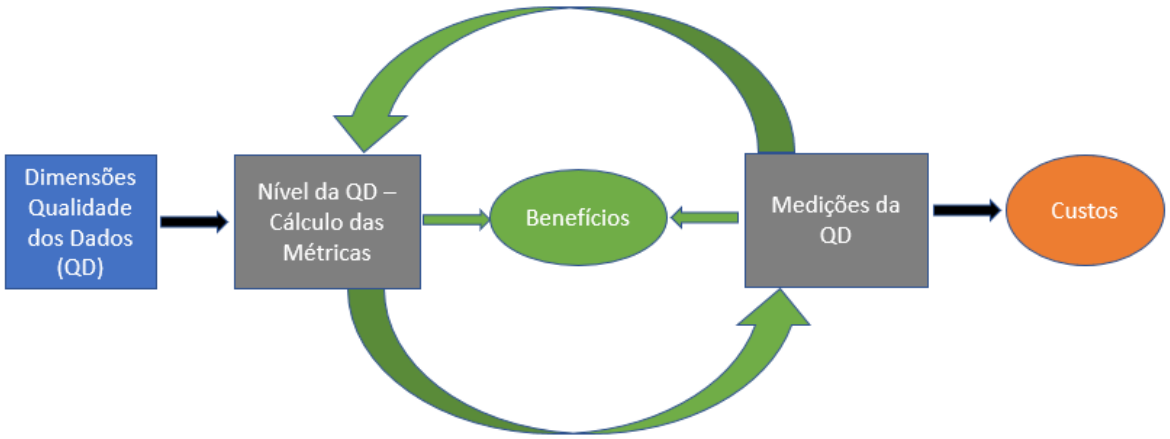

Figura 2.3 - Visualização do circuito fechado de uma gestão de qualidade de dados orientada com uma vertente económica, adaptada do estudo de (Heinrich et al., 2011)

Na prática, muitas métricas de qualidade de dados são desenvolvidas de uma forma 'ad hoc' para resolver problemas específicos e, portanto, são frequentemente afetadas por um alto nível de subjetividade. (Cappiello, Francalanci, & Pernici, 2004; Heinrich et al., 2011; Heinrich, Klier, & Kaiser, 2009; Pipino et al., 2002)

De forma a permitir uma base científica, uma boa definição e avaliação das métricas, o estudo de (Even & Shankaranarayanan, 2007; Heinrich et al., 2011; Pipino et al., 2002) apurou seis requisitos:

- Normalização - para garantir que os valores das métricas de qualidade de dados são comparáveis, é necessária uma normalização adequada (por exemplo, para permitir comparar diferentes níveis de qualidade de dados ao longo do tempo). Neste contexto, os valores dessas métricas são apurados muitas vezes com valores que variam entre 0 (muito mau) e 1 (muito bom).

- Escalas Intervalares - de modo a suportar a monitorização de como o nível da qualidade dos dados muda ao longo do tempo e a avaliação económica das medidas, as métricas devem ser ter escalas intervalares. Por exemplo, uma diferença idêntica de 0.2 entre os valores 0.7 e 0.9 e os valores 0.4 e 0.6 na medição de uma métrica para apurar a correção dos dados, significa que a quantidade de dados que estão corretos muda na mesma medida em ambos os casos.

- Interpretabilidade - as métricas de qualidade dos dados devem ser compreensíveis, logo a sua medição deve ser de fácil interpretação por parte dos utilizadores. Por exemplo, uma

métrica para apurar a atualidade dos dados pode ser interpretável como a probabilidade de um determinado valor de um atributo dentro de uma base de dados ainda estar atualizado.

- Agregação - no caso de um modelo de dados relacional, deve ser possível quantificar a qualidade dos dados a nível valores de atributos, tuplos, relações e de toda a base de dados, de forma a que os valores tenham uma interpretação semântica consistente em cada nível. Além disso, as métricas devem permitir a agregação dos valores quantificados num determinado nível para obter uma métrica para o nível seguinte. Por exemplo, a quantificação da correção de uma relação deve ser calculada com base nos valores da correção dos tuplos que fazem parte da relação.

- Adaptabilidade - para quantificar a qualidade dos dados de uma forma orientada para atingir um objetivo, é necessário que as métricas possam ser adaptadas ao contexto de uma aplicação específica. As métricas que não foram adaptadas devem registadas como medidas imparciais.

- Viabilidade - de forma a garantir a viabilidade da sua aplicação, as métricas devem ser baseadas em parâmetros de entrada que são determináveis. No momento da definição das métricas, devem ser definidos os métodos para determinar os parâmetros de entrada. Se a determinação exata dos parâmetros de entrada não for possível ou tiver custos elevados, devem ser propostos métodos alternativos. De um ponto de vista económico, também é necessário que a quantificação da qualidade dos dados possa ser realizada através de um alto nível de automação.

As métricas podem ser classificadas também como sendo objetivas, quando são baseadas em métricas quantitativas, que são aquelas que são quantificadas ou para as quais pode ser calculado um valor concreto (pontuação), ou subjetivas, que são aquelas que não podem ser quantificadas e dependem da perceção da métrica por parte dos utilizadores. Neste trabalho de projeto, irá ser dado um maior enfâse no que toca aos indicadores quantitativos, uma vez que as avaliações baseadas em medidas qualitativas, tais como questionários ou opiniões de especialistas podem ser algo subjetivas ou inconsistentes. (Vetrò et al., 2016; Zaveri et al., 2012)

2.2.6. Tipos de Dados

Os dados das organizações são distribuídos cada vez mais por diversos recursos heterogéneos e com diversos formatos diferentes, mesmo que se refiram às mesmas entidades. Esses formatos vão desde um formato de dados quase não estruturados (por exemplo, sistemas de ficheiros, repositórios de documentos e em portais web) até um formato de dados altamente estruturado (por exemplo, sistemas de gestão de bases de dados). Na literatura, as fontes de dados são classificadas de acordo com o nível de estrutura que as caracteriza. (Carlo et al., 2011)

Segundo o estudo de (Batini et al., 2009; Carlo et al., 2011), o objetivo final de uma metodologia de qualidade de dados é a análise de dados que, em geral, descrevem objetos do mundo real num formato que pode ser armazenado, recuperado e processado através da utilização de software e comunicado através de uma rede. No campo da qualidade dos dados, a literatura disponível distingue, de uma forma implícita ou explícita, três tipos de dados:

1. Dados estruturados – são agregações ou generalizações de itens descritos por atributos elementares definidos dentro de um determinado domínio. Os domínios representam o intervalo de valores que podem ser assignados a atributos, e geralmente correspondem a tipos elementares de linguagens de programação, como valores numéricos ou campos de texto. As tabelas relacionais e os dados estatísticos representam o tipo de dados estruturados mais comum.

2. Dados não estruturados – é uma sequência genérica de símbolos, tipicamente representada através de uma linguagem natural. Exemplos típicos de dados não estruturados são questionários que contêm campos de texto livre para a resposta a perguntas abertas, ou o corpo de um email. Nestes casos não existe uma estrutura específica definida, nem são definidos domínios dos tipos de dados nem restrições formais.

3. Dados semi-estruturados – dados que possuem uma estrutura que tem algum grau de flexibilidade. São dados semi-estruturados se os dados estão situados entre dados em bruto e os dados que foram inseridos pelo utilizador, ou seja, quando têm alguma estrutura, mas não têm uma estrutura tão rígida como a que existe nas bases de dados e são geralmente representados por linguagem XML. Algumas características comuns são: (1) os dados podem conter campos que não são conhecidos na altura do desenho da estrutura (por exemplo, um ficheiro XML pode não conter um esquema XML associado); (2) o mesmo tipo de dados pode ser representado de várias maneiras (por exemplo, uma data pode ser representada por um campo ou por vários campos, até mesmo dentro do mesmo conjunto de dados; (3) entre os campos conhecidos na altura do desenho da estrutura, muitos não terão valores.

Podemos verificar um exemplo prático, retirado do estudo de (Batini et al., 2009), com diferentes representações do mesmo objeto do mundo real. Neste exemplo será considerado um registo que contém e descreve informações pessoais, tais como o Nome, Sobrenome, Região e País de Nascimento. Podemos então verificar neste exemplo a representação do Sr. Patrick Metzisi, nascido na região de Masai Mara, no Quénia:

a) Dados Estruturados

Patrick	Metzisi	Masai Mara	Quénia

b) Dados não estruturados

Sr. Patrick Metzisi, nascido na região de Masai Mara, no Quénia

c) Dados semi-estruturados

```
<DadosPessoais>
    <nome>Patrick Metzisi</nome>
    <naturalidade>Masai Mara, Quénia</naturalidade>
```

```
</DadosPessoais>
```

As diferenças no formato dos dados são necessariamente refletidas nos métodos e técnicas que as organizações utilizam para avaliar e melhorar a qualidade dos seus recursos de informação. Essas diferenças levam a uma utilização de metodologias que apoiam a seleção e implementação de programas de melhoria da qualidade dos dados que são adaptados a necessidades e domínios específicos. Por exemplo, a manutenção das grandes bases de dados é efetuada através da utilização de técnicas de limpeza e técnicas de comparação de registos, ao passo que a qualidade dos documentos de uma organização é melhorada através da adoção de modelos mais estruturados e léxicos mais formais e inequívocos e estabelecendo procedimentos internos de auditoria. (Carlo et al., 2011)

A identificação das dimensões e métricas da Qualidade dos Dados varia com o tipo de dados. Para dados estruturados e semi-estruturados, a qualidade dos dados é geralmente medida através das dimensões de qualidade como *Accuracy*, *Completeness* e *Currency*, dado que são dimensões independentes do contexto e associadas a algoritmos de avaliação consolidados. (Batini et al., 2006)

Para dados não estruturados, as técnicas de avaliação são menos consolidadas. Por exemplo, é bastante difícil avaliar dimensões como *Accuracy* e *Completeness* num texto. A avaliação de uma dimensão como *Currency* é mais fácil e é mais comum, uma vez que a frequência de atualização pode ser facilmente medida para um texto. (Batini et al., 2006)

Na Tabela 7.2 da secção de Apêndices, pode ser consultado o estudo do tipo de dados de cada conjunto de dados que se encontram disponibilizados no Portal de Dados Abertos da cidade de Lisboa.

2.3. *DASHBOARDS*

A visualização dos dados está a ganhar um interesse cada vez maior, como uma ferramenta indispensável para a exploração e análise dos dados num conjunto diversificado de aplicações orientadas para o efeito. Através destas aplicações, os analistas de dados exploram grandes volumes de dados, através de visualizações que revelam novas e valiosas descobertas. (Ehsan, Sharaf, & Chrysanthis, 2016)

Um processo de avaliação da qualidade dos dados tem como umas das suas premissas principais que um defeito na qualidade dos dados denota uma não conformidade entre uma instância de dados e o seu significado contextual, que pode surgir em qualquer ponto do ciclo de vida dos dados. Existem diversas abordagens que suportam esse processo que aplicam métodos quantitativos que geralmente limitam a interpretação humana dos seus resultados. (Borovina Josko & Ferreira, 2017)

No entanto, um processo de avaliação da qualidade dos dados depende fortemente do conhecimento do contexto dos dados, uma vez que é o contexto que determina a estrutura de significado entre os dados e o ambiente em que são utilizados. Neste sentido, a supervisão humana é essencial em todo este processo e os sistemas de visualização pertencem a uma classe de

abordagens supervisionadas que combinam a capacidade computacional com a descoberta de padrões e distinções semânticas inatas para os seres humanos, de modo a permitir a avaliação visual da qualidade dos dados. (Borovina Josko & Ferreira, 2017)

Através de representações visuais interativas, quem avalia a qualidade dos dados consegue pesquisar, extrair, correlacionar e entender os significados (padrões, relacionamentos e métricas) em diferentes granularidades até serem indicadas evidências semânticas que confirmem ou refutem um defeito nos dados. (Borovina Josko & Ferreira, 2017)

Os sistemas de visualização projetados para suportar um processo de análise da qualidade dos dados baseiam-se em duas abordagens. A abordagem denominada como reconhecimento de qualidade, que denota a utilização de recursos computacionais para extrair métricas de qualidade dos dados que são visualmente comunicadas através de *highlights* ou rótulos. A abordagem denominada como orientação para o diagnóstico visual, que denota uma análise visual intensiva de significados para detetar defeitos de dados. (Borovina Josko & Ferreira, 2017)

Neste projeto irá ser desenvolvido um sistema de visualização, através do desenvolvimento e implementação de um *mockup* de um *Dashboard* que suporte um processo de análise da qualidade dos dados apoiada na abordagem denominada como reconhecimento de qualidade, que denota a utilização de recursos computacionais para extrair métricas de qualidade dos dados que são visualmente comunicadas através de *highlights* ou rótulos.

Nenhum exemplo de visualização de dados ocupa um lugar mais proeminente na consciência das pessoas de negócios que um *dashboard*. O display de um *Dashboard* combina toda a informação que é necessária para monitorizar rapidamente o cenário de um determinado aspeto do negócio num único ecrã. Quando são projetados de forma apropriada para uma comunicação visual eficaz, os *dashboards* oferecem uma imagem do que está a ocorrer, o que nunca poderia ser efetuado através de relatórios tradicionais. (Few, 2007)

2.3.1. Características do desenho de um *Dashboard*

Segundo o estudo de (Yigitbasioglu & Velcu, 2012), as características do desenho de um *Dashboard* podem ser divididas em dois tipos, as características funcionais e as características visuais.

As características funcionais são as características que se relacionam indiretamente com a visualização, mas que descrevem o que um *Dashboard* pode fazer. É importante que as características funcionais de um *Dashboard* se ajustem aos seus objetivos, uma vez que um ajuste inadequado pode resultar em resultados que não serão os mais perfeitos, fornecendo indicações incompletas a quem necessita do *Dashboard* para tomar as melhores decisões. As características funcionais permitem o ajuste cognitivo com os diferentes tipos de utilizadores de um *Dashboard* e são compostas pelos seguintes pontos:

- Tipos de formato da apresentação (Gráficos vs. Tabelas)
- Flexibilidade no formato da apresentação
- *Drill down and drill up*
- Análise do cenário

- Seleção do formato da apresentação guiada pela teoria
- Alertas automáticos

No entanto, mesmo que as características funcionais de um *Dashboard* estejam bem ajustadas aos objetivos (ou seja, que todas as informações e recursos necessários estão disponíveis ao utilizador), uma má definição das características visuais (por exemplo, a utilização excessiva de cores, uma baixa taxa de utilização de cores na representação dos dados, etc.) pode confundir e distrair o utilizador do *Dashboard*. As características visuais de um *Dashboard* potenciam uma melhor visualização da informação e são compostas pelos seguintes pontos:

- Página única
- Uso frugal de cores
- Alta taxa de utilização de cores na representação dos dados
- Utilização de linhas de grelha em gráficos 2D e 3D

2.3.2. Utilizadores alvo do *Dashboard*

Um dos aspetos mais importantes, se não o mais importante, é a apresentação no *Dashboard* dos indicadores que realmente importam para os utilizadores alvo, tendo em conta os objetivos e o problema que foi definido.

A identificação das variáveis cruciais associadas a um *Dashboard* tem uma enorme importância e a seleção de dados adequados para a eficácia do sistema de medição baseado num *Dashboard* também é um aspeto fundamental. (Marco, Mangano, & Zenezini, 2015)

Segundo o estudo de (Juice, 2009), os *Dashboards* são úteis porque aquilo que é medido é melhorado e porque é muito importante um entendimento compartilhado do estado do negócio.

No contexto deste projeto é importante determinar os motivos específicos pelos quais a utilização de um *Dashboard* irá ser útil para a resolução do problema e o estudo de (Juice, 2009) indica três questões chaves a que se deve dar resposta:

1. Quem é o utilizador alvo do *Dashboard*?
    - A Equipa de Gestão do Portal de Dados Abertos da Cidade de Lisboa
2. Qual o valor que o *Dashboard* irá adicionar?
    - Definir metas e expectativas para a equipa de Gestão da Qualidade dos Dados relacionadas com as diversas dimensões e características da qualidade dos dados que são inseridos no Portal de Dados Abertos da Cidade de Lisboa
    - Encorajar ações específicas sobre os conjuntos de dados antes de serem colocados no Portal de Dados Abertos de Lisboa, de modo a promover uma maior qualidade nos dados que ali serão colocados

- Realçar exceções e fornecer alertas quando existirem problemas na qualidade dos dados que irão ser colocados no Portal de Dados Abertos da Cidade de Lisboa

- Comunicar o progresso e o sucesso da aplicação das diferentes metodologias da qualidade dos dados nos dados que serão colocados no Portal de Dados Abertos da Cidade de Lisboa

- Fornecer uma interface comum para interagir e analisar os resultados da aplicação das diferentes metodologias de qualidade dos dados aos dados que irão ser disponibilizados no Portal de Dados Abertos da cidade de Lisboa

3. Qual o tipo de *Dashboard* deve ser criado?

    - Um tipo de *Dashboard* focado no âmbito da aplicação de metodologias de qualidade dos dados

    - Fornecendo uma visão focada na qualidade dos dados que irão ser colocados no Portal de Dados Abertos da Cidade de Lisboa

    - Com um horizonte temporal histórico, de modo a ir verificando a evolução da qualidade dos dados ao longo do tempo, depois de aplicadas algumas correções

    - Apresentado com uma visão única para todos os utilizadores, com vários níveis de detalhe, proporcionando a capacidade de o utilizador efetuar *drill down*, para ter acesso a números mais detalhados para ganhar mais contexto

    - Com um ponto de vista exploratório, em que o utilizador tem a liberdade de interpretar os resultados conforme achar mais apropriado

Os *Dashboards* bem-sucedidos são aqueles que contribuem para melhorar a tomada de decisões e o desempenho organizacional através da utilização dos dados certos no momento certo, para tomar as decisões mais acertadas. (Anandarajan & Jones, 2017)

2.4. PROPOSTA CONCEPTUAL DO MODELO DE AVALIAÇÃO DA QUALIDADE DOS DADOS

O modelo de avaliação da qualidade dos dados que irá ser implementado neste projeto tem como base o estudo na literatura de um modelo de dados que permita implementar uma *framework* para a avaliação e tratamento da qualidade dos dados que irão ser colocados no Portal de Dados Abertos da cidade de Lisboa. Esse modelo de dados deve permitir a aplicação de diferentes metodologias de tratamento da qualidade dos dados, já estudadas no capítulo 2.2.4.

Esta proposta de modelo de avaliação da qualidade dos dados e que foi a base para o desenvolvimento do projeto encontra-se descrita no capítulo 3.1.

No seguimento da implementação do modelo de avaliação da qualidade dos dados que irão ser colocados no Portal de Dados Abertos da cidade de Lisboa, irá ser desenvolvido um *mockup* de um *Dashboard* de modo a apresentar aos utilizadores os resultados do funcionamento da *framework* de tratamento da qualidade dos dados que irão ser colocados no Portal de Dados Abertos da cidade de

Lisboa. A proposta de protótipo do *Dashboard* encontra-se descrita no ponto 2.4.1. através de um *wireframe*.

2.4.1. Wireframe do Dashboard

A utilização de protótipos para a visualização de interfaces (geralmente conhecidas como *mockups* ou *wireframes*), provaram aumentar a eficiência na perceção dos requisitos das aplicações. Uma das suas vantagens prende-se com o facto dos *wireframes* serem tecnicamente valiosos para os programadores e, ao mesmo tempo, serem totalmente compreensíveis para os utilizadores finais. (Rivero et al., 2014)

A maioria dos clientes preocupa-se principalmente como vão interagir com o produto e preocupa-se menos com os detalhes. Para terem uma ideia clara do produto têm de visualizar o mesmo e isso pode ser alcançado através do desenvolvimento de um *mockup* do produto antes do desenvolvimento das especificações. Este tipo de *mockup* é muitas vezes chamado de *wireframe*, é barato de criar e uma ferramenta de comunicação muito eficaz. (Kreitzberg, 2004)

Criar um *wireframe* no processo inicial de criação da interface das aplicações com o utilizador é muito útil como meio de comunicação entre os utilizadores e quem desenvolve o produto.

O *wireframe* serve como guia visual que é utilizado para sugerir o conteúdo e a estrutura do produto, assim como as interações e as relações na utilização do mesmo, que serão afinadas até que exista acordo com os utilizadores.

Assim sendo, no âmbito deste projeto foi desenvolvido um *wireframe* do *Dashboard* que irá servir para apresentar aos utilizadores os resultados do funcionamento da *framework* de tratamento da qualidade dos dados que irão ser colocados no Portal de Dados Abertos da cidade de Lisboa, que pode ser verificado na figura 2.4.

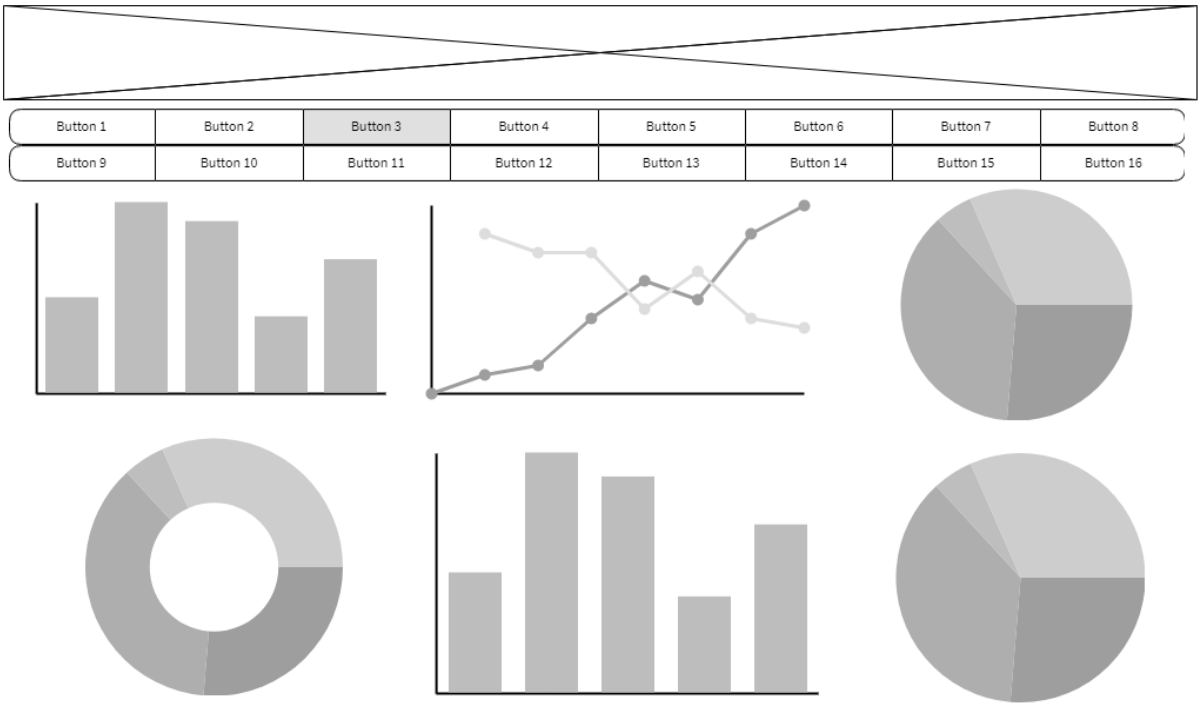

Figura 2.4 - *Wireframe* do *Dashboard* com os indicadores da Qualidade dos Dados que foram avaliados na *framework* de avaliação da qualidade dos dados

O *wireframe* do *Dashboard* que irá ser desenvolvido e que se encontra representado na figura 2.4, contém:

- Uma imagem no topo que irá conter o título do *Dashboard*
- Um conjunto de botões que irá ter a função de *slicer* e cada botão deverá representar um grupo de dados daqueles que compõem o Portal de Dados Abertos da cidade de Lisboa.
- Um conjunto de gráficos que irão representar as métricas de cada dimensão que foram calculadas para indicar os níveis de qualidade dos dados que irão ser colocados no Portal de Dados Abertos da cidade de Lisboa. Através dos botões que têm a função de *slicer*, o utilizador poderá ir verificando os indicadores de qualidade de cada dimensão que foram calculados em cada grupo de dados. Inicialmente serão apresentados os dados gerais, sem qualquer tipo de filtro.

Este *wireframe* foi elaborado com recurso a uma ferramenta de construção de *wireframes*, denominada como 'Mockflow' e está disponível através do site (https://www.mockflow.com/).

3. DESENVOLVIMENTO

Como já foi mencionado na definição da metodologia, o desenvolvimento deste projeto tem como base um processo cíclico, o que significa que é um método iterativo usado para um controlo contínuo e melhoria dos artefactos. Para iniciar o processo, uma sugestão inicial deve ser feita. Isso conduzirá a uma fase de implementação e, finalmente, a uma fase de avaliação dos resultados. A avaliação determinará o início de um novo ciclo ou à conclusão do projeto.

A fase inicial deste processo de desenvolvimento prende-se com o estudo na literatura de um modelo de dados que permita implementar uma *framework* para a avaliação e tratamento da qualidade dos dados que irão ser colocados no Portal de Dados Abertos da cidade de Lisboa. Esse modelo de dados deve permitir a aplicação de diferentes metodologias de tratamento da qualidade dos dados, já estudadas no capítulo 2.2.4.

Numa fase seguinte, irá ser iniciada uma fase de implementação da *framework* para a avaliação e tratamento da qualidade dos dados que foi estudada e será desenvolvido um protótipo inicial para verificar o comportamento do funcionamento da *framework* e os seus indicadores.

Este capítulo irá terminar com o desenvolvimento de um *mockup* de um *Dashboard* de modo a apresentar aos utilizadores os resultados do funcionamento da *framework* de tratamento da qualidade dos dados que irão ser colocados no Portal de Dados Abertos da cidade de Lisboa.

3.1. FRAMEWORK PARA A AVALIAÇÃO DA QUALIDADE DOS DADOS

Os dados abertos são uma parte do conteúdo ou dos dados, com o intuito de alguém poder utilizá-los, reutilizá-los e também redistribuí-los. A maioria dos conjuntos de dados abertos divulgados está ainda em formato bruto e o seu sucesso depende fortemente da sua qualidade. (Máchová & Lněnička, 2017)

A difusão dos Dados Abertos manteve um ritmo muito elevado nos últimos anos. No entanto, existem evidências que indicam que a divulgação de dados sem o controlo de qualidade adequado pode comprometer a reutilização dos conjuntos de dados e afetar negativamente a participação cívica. (Vetrò et al., 2016)

Formalmente, a Qualidade dos Dados é definida na Norma ISO 25012 como "a capacidade de dados para satisfazer necessidades declaradas e implícitas quando usadas sob condições especificadas". Além dessa definição, a Qualidade dos Dados pode ser definida como a capacidade de uma coleta de dados que vão de encontro às necessidades dos utilizadores, ou seja, que tenham uma "adequação ao uso" tendo em conta o ponto de vista do utilizador sobre a qualidade que esses dados devem ter. (Batini et al., 2009; ISO/IEC 25012, 2008; Vetrò et al., 2016; Wang & Strong, 1996)

É neste contexto que o estudo de (Batini et al., 2007) apresenta uma *framework* para a avaliação da qualidade dos Dados, que no caso concreto deste projeto, serão os dados que irão ser disponibilizados no Portal de Dados Abertos da cidade de Lisboa.

A arquitetura desta *framework* (Figura 3.1) é composta por cinco módulos (e respetivos repositórios associados), sendo que o desenvolvimento deste projeto irá ter em conta os seguintes quatro módulos:

- *Knowledge Extractor* – O módulo '*Knowledge Extractor*' permite a definição de todos os relacionamentos entre os dados utilizados nos processos, serviços e pelas unidades de negócio, armazenando-os num repositório.
- *Data Quality Assessment* - O módulo '*Data Quality Assessment*' é responsável por avaliar a qualidade das fontes de dados, aplicando diferentes algoritmos e técnicas, armazenando todos os resultados num repositório.
- *Analysis* – O módulo '*Analysis*' tem como objetivo o processamento e análise da informação retirada do módulo de '*Data Quality Assessment*', sendo que a análise OLAP é armazenada num repositório.
- *Monitoring & Reporting* – O módulo '*Monitoring & Reporting*' permite a realização de atividades de monitorização e de relatórios sobre as informações mais importantes.

Figura 3.1 – Arquitetura da *framework* de avaliação da Qualidade dos Dados, adaptada do estudo de (Batini et al., 2007)

3.1.1. Arquitectura da framework

Com base na *framework* de avaliação da Qualidade dos Dados apresentada no ponto anterior, irá ser implementado um sistema de *Data Warehouse* (DW) com uma arquitetura do tipo Star (Figura 3.2.), onde existirão tabelas factuais ligadas às suas respetivas dimensões através de *Surrogate Keys*. Serão criados *Jobs* de *extract-transform-load* (ETL), responsáveis pela extração, tratamento/transformação e posterior carregamento nas tabelas do DW.

As Dimensões são tabelas que armazenam os conceitos de negócio, os quais irão caracterizar os registos das tabelas factuais. As dimensões implementadas no DW de Gestão da Qualidade dos dados serão do tipo zero, um e dois. As dimensões do tipo zero, são normalmente carregadas uma vez e não sofrem mais alterações dos seus registos. As dimensões do tipo um, podem sofrer alterações a alguns os seus campos, sendo o sistema capaz de efetuar atualizações a esses registos.

Por último as dimensões do tipo dois, são idênticas às do tipo um, mas permitem reter histórico das alterações dos seus campos, através de um campo de '*start date*' e '*end date*', o qual representa o período temporal em que determinado registo é válido. Os registos atuais são caracterizados por possuírem o campo '*end date*' com o valor 'NULL'.

As tabelas factuais servem o propósito de armazenar os dados transacionais diários, possuindo diversos campos com métricas úteis para a avaliação da qualidade dos dados. Cada uma das transações será sempre caracterizada pela dimensão tempo e por um conjunto de chaves de outras dimensões que caracterizam a relação.

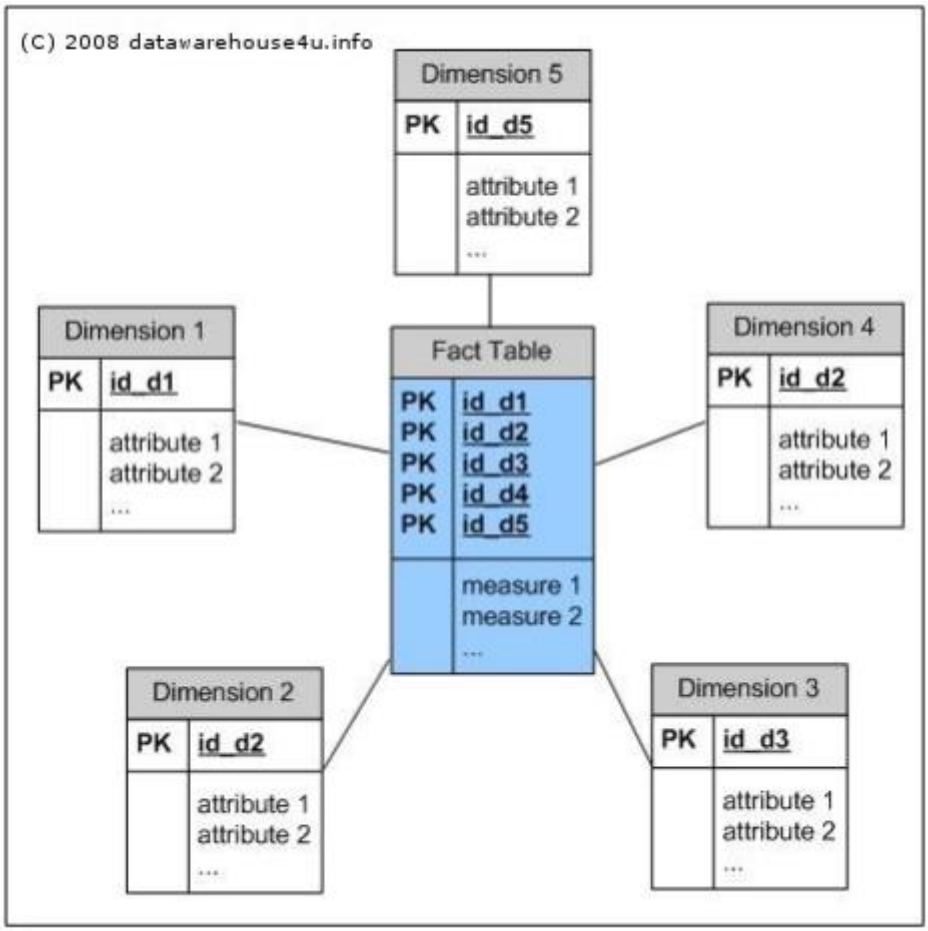

Figura 3.2 – Exemplo de arquitetura do tipo Star, retirada de (www.datawarehouse4u.info, n.d.)

Numa primeira fase, irá ser efetuado o levantamento e identificação das fontes de informação do sistema operacional, que no caso deste Projeto se encontram todas no Portal de Dados Abertos da Cidade de Lisboa. Este é maioritariamente constituído por ficheiros externos em formato CSV, GeoJSON, JSON ou PDF.

Numa segunda fase, e seguindo o modelo da *framework* indicada no estudo de (Batini et al., 2007) será implementado um módulo designado como '*Knowledge Extractor*', que irá conter a definição de todos os relacionamentos entre os dados utilizados nos processos, serviços e pelas unidades de negócio, armazenando-os num repositório.

Numa terceira fase, irá ser desenvolvida a definição dos processos ETL para obter os dados da BD operacional e ficheiros fonte, carregando estes dados numa base de dados denominada por *Staging Area*. Esta área tem como objetivo melhorar a performance da fase de ETL, sendo os dados carregados dos sistemas fontes de uma forma ainda pouco transformada. As tabelas da *Staging Area* são igualmente eficientes pois não efetuam armazenamento de dados históricos, sendo que os dados contidos nestas tabelas são apagados no início de cada execução, permitindo carregamentos mais céleres e consultas mais eficientes na fase de carregamento das tabelas DW. Esta área reflete também o módulo de '*Data Quality Assessment*' indicado na *framework* do estudo de (Batini et al., 2007).

Numa quarta fase, são efetuadas as transformações finais aos dados existentes na *Staging Area* e são feitos os carregamentos das tabelas de Dimensões e tabelas Factuais. O DW terá cerca de 3 dimensões e 1 tabela fatual. Esta área reflete o módulo de '*Analysis*' indicado na *framework* do estudo de (Batini et al., 2007).

Na figura 3.3. é possível verificar o desenho da Arquitetura do DW para análise da qualidade dos dados que serão inseridos no Portal de Dados Abertos da cidade de Lisboa.

Na fase final, será implementado um *mockup* de um *Dashboard*, de modo a que o utilizador final possa realizar atividades de monitorização sobre as informações mais importantes, que neste caso concreto serão as métricas apuradas sobre a qualidade dos dados a serem colocados no Portal de Dados Abertos da cidade de Lisboa. Esta área reflete o módulo de '*Monitoring & Reporting'* indicado na *framework* do estudo de (Batini et al., 2007).

As ferramentas que irão ser utilizadas na elaboração do DW será o '*SQL Server Management Studio 2016*' (SSMS). Os processos de ETL serão definidos em '*SQL Server Integration Services*' (SSIS) através da ferramenta de desenvolvimento '*SQL Server Data Tools 2015*' e o desenvolvimento do *mockup* do *Dashboard* irá ser efetuado com recurso à ferramenta '*Power BI*'.

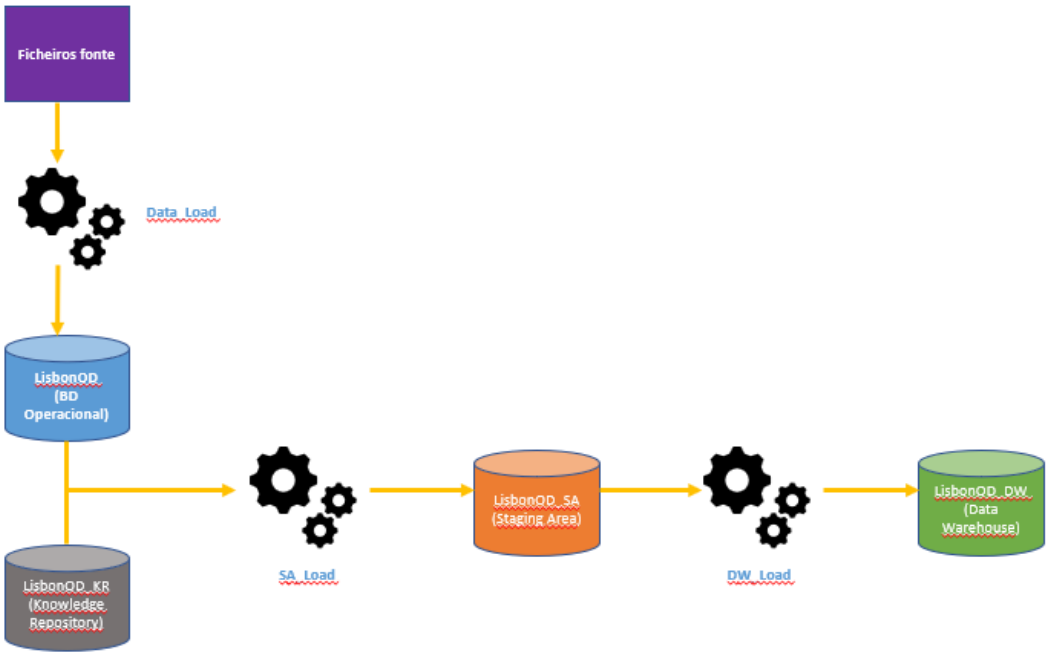

Figura 3.3 – Desenho da Arquitetura do DW para análise da qualidade dos dados que serão inseridos no Portal de Dados Abertos da cidade de Lisboa

As tabelas de DW seguiram o modelo de relações que se pode verificar na figura 3.4:

- Factual de Métricas

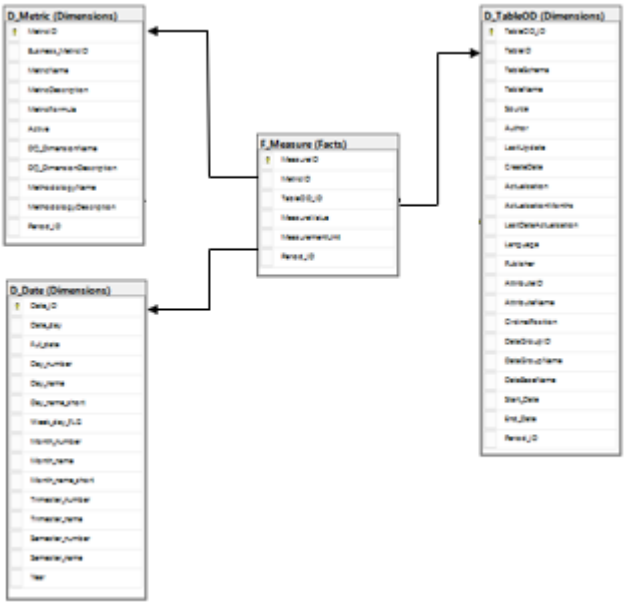

Figura 3.4 – Modelo de dados do DW para a avaliação da qualidade dos dados do Portal de Dados Abertos da cidade de Lisboa

3.1.2. Fontes de Dados

Para o desenvolvimento deste Projeto, os ficheiros fonte são todos ficheiros que se encontram disponíveis no Portal da Dados Abertos da cidade de Lisboa. Foi efetuado um levantamento e um estudo de todos os ficheiros disponíveis, verificando qual o tipo de ficheiro e se continham dados Estruturados, Semi-Estruturados ou Não Estruturados (Tabela 7.2. do capitulo de Apêndices).

De acordo com esse levantamento, no Portal estão presentes ficheiros com os seguintes formatos: CSV, XLS, XLSX, PDF, JSON, GeoJSON, SHP, GTFS, TIFF e PHP.

Assim sendo, foi decidido retirar uma amostra de cada grupo de dados (Tabela 7.3. do capítulo de Apêndices), para a avaliação da sua qualidade. Esses grupos de dados estão indicados no capítulo 2.2.1. deste documento.

Foi então efetuado o carregamento dessa amostra de dados para uma base de dados operacional, designada como 'LisbonOD', uma vez que os vários tipos de dados necessitavam de um tipo de processamento diferente de uns para os outros e foi decidido centralizar tudo numa base de dados, de modo a prosseguir com o processo.

A descrição das tabelas operacionais que compõem a base de dados 'LisbonOD', encontra-se na Tabela 7.3. do capítulo de Apêndices.

3.2. MÓDULO 'KNOWLEDGE EXTRACTOR'

Seguindo o modelo da *framework* indicada no estudo de (Batini et al., 2007) foi implementado um módulo designado como '*Knowledge Extractor*', que contém a definição de todos os relacionamentos entre os dados utilizados nos processos, serviços e pelas unidades de negócio, armazenando-os numa base de dados, designada como 'LisbonOD_KR'.

Na tabela 3.1, podemos verificar a composição da base de dados 'LisbonOD_KR' e a descrição das tabelas que a compõem.

Tabela 3.1 – Descrição das tabelas que compõem a base de dados 'LisbonOD_KR', do módulo '*Knowledge Extractor*'

Tabela	Descrição
Know_Repository.LisbonOD_DataGroup	Tabela que contém a descrição de todos os grupos de dados do Portal de Dados Abertos da cidade de Lisboa
Know_Repository.LisbonOD_DataBase	Tabela que contém a descrição de todas as bases de dados do Portal de Dados Abertos da cidade de Lisboa
Know_Repository.LisbonOD_Table	Tabela que contém a descrição de todas as tabelas que compõem as bases de dados do Portal de Dados Abertos da cidade de Lisboa

Know_Repository.TableAttribute	Tabela que contém a descrição de todos os atributos das tabelas que compõem as bases de dados do Portal de Dados Abertos da cidade de Lisboa
Know_Repository.Methodology	Tabela que contém a descrição das metodologias que compõem o processo de medição da qualidade dos dados
Know_Repository.DQ_Dimension	Tabela que contém a descrição de todas as dimensões/características que compõem o processo de medição da qualidade dos dados
Know_Repository.Metric	Tabela que contém a descrição de todas as métricas que compõem o processo de medição da qualidade dos dados

3.3. MÓDULO 'STAGING AREA'

Seguindo o modelo da *framework* indicada no estudo de (Batini et al., 2007) foi implementado um módulo designado como *Staging Area (SA)*, que reflete o módulo de '*Data Quality Assessment*' dessa mesma *framework*. É neste módulo que serão efetuadas algumas operações de modo a avaliar a qualidade das fontes de dados, aplicando os diferentes algoritmos e técnicas que compõem as métricas e armazenando todos os resultados num repositório.

As tabelas desta área serão limpas na sua totalidade (*Truncate*) no início de cada processamento. Todas as tabelas da SA terão um campo denominado 'Period_ID'. Este valor corresponde á data em que é efetuado o carregamento dos dados para as tabelas. Desta forma é mais simples fazer o seguimento de dados carregados.

Na tabela 3.2, podemos verificar a composição da base de dados 'LisbonOD_SA' e a descrição das tabelas que a compõem. Na figura 3.5, podemos também verificar o modelo de dados deste módulo de SA.

Tabela 3.2 - Descrição das tabelas que compõem a base de dados 'LisbonOD_SA', do módulo '*Staging Area*'

Tabela	Descrição
Staging_Area.SA_Metric	Tabela que contém a descrição de todas as métricas que compõem o processo de medição da qualidade dos dados
Staging_Area.SA_TableOD	Tabela que contém a descrição de todas as tabelas, os seus atributos e a base de dados e o grupo de dados a que pertencem
Staging_Area.Measurement	Tabela que contém todas as medições que resultaram da aplicação de diferentes algoritmos e técnicas que compõem as métricas que foram aplicadas aos conjuntos de dados do Portal de Dados Abertos da cidade de Lisboa

A descrição de todos os campos das tabelas indicadas na Tabela 3.2, pode ser verificada nas Tabelas 7.4, 7.5 e 7.6, no capítulo de Apêndices.

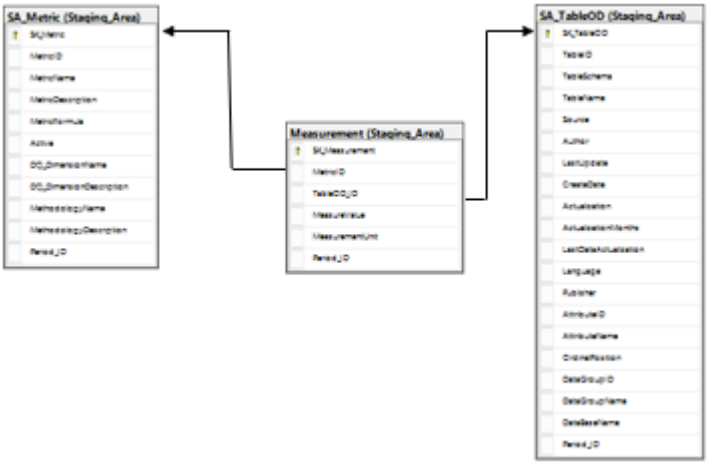

Figura 3.5 - Modelo de dados do módulo de *Staging Area*

3.4. MÓDULO 'DATA WAREHOUSE'

Seguindo o modelo da *framework* indicada no estudo de (Batini et al., 2007) foi implementado um módulo designado como *Data Warehouse (DW)*, que reflete o módulo de '*Analysis*' dessa mesma *framework*. Este módulo que tem como objetivo o processamento e análise da informação retirada do módulo de *Staging Area.*

Foi criado um *schema* 'Dimensions', onde estarão criadas as dimensões do sistema. O outro *schema* criado será o 'Facts', onde estará a tabela fatual representada.

Na tabela 3.3, podemos verificar a composição da base de dados 'LisbonOD_DW' e a descrição das tabelas que a compõem. Na figura 3.4, podemos também verificar o modelo de dados deste módulo de DW.

Tabela 3.3 - Descrição das tabelas que compõem a base de dados 'LisbonOD_DW', do módulo '*Data Warehouse*'

Tabela	Descrição
Dimensions.D_Date	Tabela com informação e hierarquias de tempo. Dado que esta dimensão é refrescada com muito pouca regularidade, não foi criada uma tabela de SA para efetuar o seu carregamento. É uma dimensão do Tipo 0.

Dimensions.D_Metric	Dimensão que guardará a informação referente às métricas que são aplicadas para apurar a qualidade dos dados. Sendo uma Dimensão do tipo 1, qualquer alteração aos campos 'Active', 'MetricFormula' e 'MetricName' terá uma operação de atualização da informação.
Dimensions.D_TableOD	Dimensão que guardará a informação referente aos dados das tabelas onde se encontram os dados cuja qualidade vai ser avaliada. Sendo uma dimensão do tipo 2, existirá um campo de 'Start Date' e 'End Date', que definem o período temporal de validade do registo. Qualquer alteração aos campos relacionados com a atualização da tabela ou dos seus dados, resultará na criação de um novo registo e de atualização da data de fecho do registo atual.
Facts.F_Measure	Tabela fatual de armazenamento dos resultados da aplicação das métricas de avaliação da qualidade dos dados que irão ser colocados no Portal de Dados Abertos. Cada registo será caracterizado por vários atributos e Id's de duas dimensões. Contém várias métricas importantes para o apuramento da avaliação da qualidade dos dados. O carregamento desta tabela será efetuado de uma forma incremental, e para ser evitada uma duplicação quando existe a necessidade de efetuar um reprocessamento, será efetuada uma operação de *delete* por 'Period_ID', procedendo-se posteriormente ao carregamento.

A descrição de todos os campos das tabelas indicadas na Tabela 3.3, pode ser verificada nas Tabelas 7.7, 7.8, 7.9 e 7.10, na secção dos Apêndices.

3.5. PROCESSOS DE ETL

Nesta secção do documento será apresentada a lógica seguida para a construção do processo de ETL do *Data Warehouse* que irá conter todos os dados das medições da avaliação da qualidade dos dados. Foi decidido separar o processo de ETL para carregamento da SA e do DW. Para cada um dos processos de carregamento existirão dois *main containers* (Dimensões e Fatuais) e para cada um desses *containers*, um *container* por fluxo. (Figura 3.6)

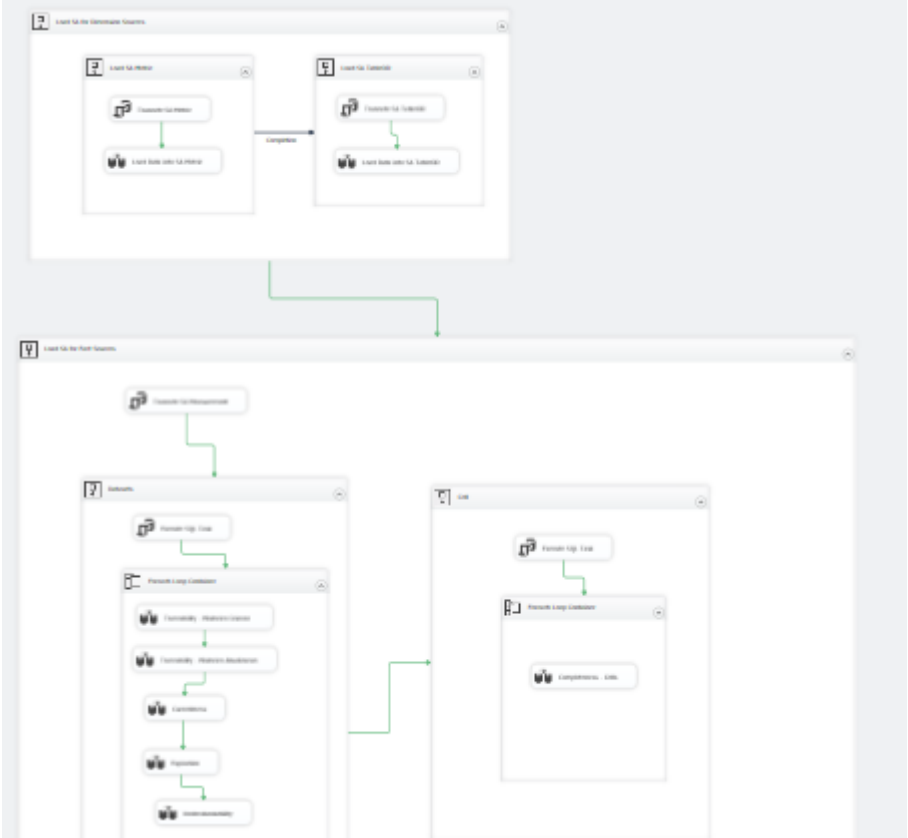

Figura 3.6 – *Main Containers* do Processo de ETL

3.5.1. Control Flow

A lógica genérica de carregamento em "Control Flow" de cada um dos fluxos será idêntica, seguindo os seguintes passos (Figura 3.7):

1. Obter o Period_ID para o fluxo em questão e mapear este valor para uma variável denominada como 'Period_ID'.
2. Proceder ao *truncate* da tabela através de uma SQL Task, no caso das tabelas de *Staging* ou efetuar um delete por Period ID no caso das tabelas fatuais incrementais.
3. Obter a listagem de TableID's e AttributeID's e mapear os valores para variáveis, de modo a poderem ser percorridos no "Data Flow" da tabela factual.
4. Evocar o "Data Flow" de carregamento dos dados.
5. Proceder ao carregamento do fluxo seguinte.

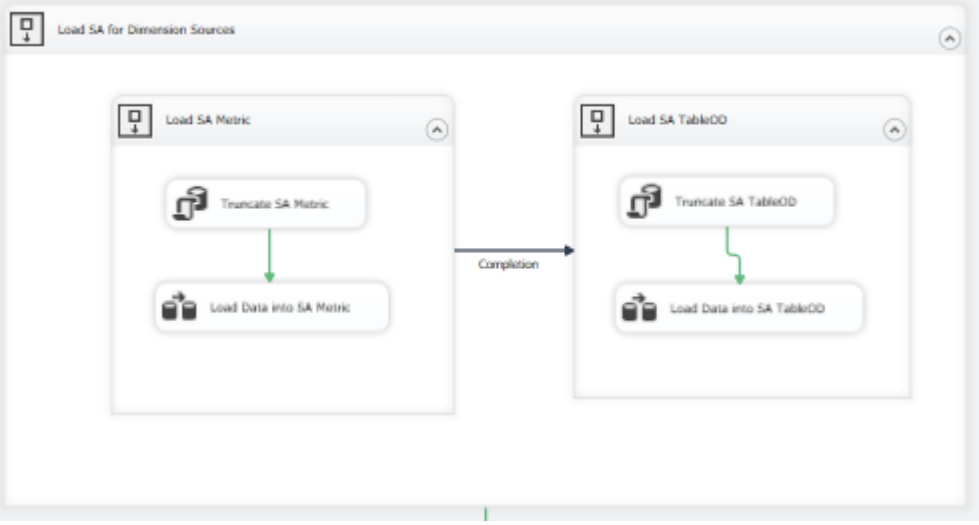

Figura 3.7 – Exemplo de carregamento por fluxos – SA *Load Dimension Sources*

3.5.1.1. Control Flow – SA

O fluxo específico de carregamento dos dados na área de "Control Flow" em contexto de SA seguiu a seguinte ordem:

1. Efetuar a operação de *Truncate* e carregamento da tabela fonte de uma das dimensões, a tabela 'SA_Metric'
2. Efetuar a operação de *Truncate* e carregamento da tabela fonte de uma das dimensões, a tabela 'SA_TableOD'
3. Efetuar a operação de *Truncate* e carregamento da tabela fonte da tabela fatual, a tabela 'SA_Measurement'

Na operação relativa ao ponto 3, foram efetuados os cálculos das métricas que se encontram indicadas na tabela 3.4. Estas métricas foram retiradas de duas das metodologias estudadas durante a elaboração deste Projeto.

Tabela 3.4 – Dimensões/Características estudadas e implementadas neste Projeto

Dimensão/ Característica	**Métrica**	**Nível de Análise**	**Descrição**	**Referência**
Traceability	Histórico da criação	Tabela	Indica a presença ou ausência de metadados associados ao processo de criação de um conjunto de dados.	(Vetrò et al., 2016)
	Histórico de atualizações	Tabela	Indica a existência ou ausência de metadados associados às atualizações	(Vetrò et al., 2016)

			feitas num conjunto de dados.	
Currentness	Indicação de versão atualizada de um Conjunto de Dados	Tabela	Indicação da atualidade da versão de um Conjunto de Dados.	(Batini et al., 2009; Vetrò et al., 2016)
Expiration	Indicação do atraso após a expiração da versão atual de um Conjunto de Dados	Tabela	Indica a relação entre o atraso na publicação de um conjunto de dados após a data de expiração da sua versão anterior e o período de tempo referido pelo conjunto de dados, em meses.	(Vetrò et al., 2016)
Completeness	Percentagem de células completas em cada coluna das Tabelas	Atributo	Indica a percentagem de células que não estão vazias em cada coluna das tabelas	(Batini et al., 2009; Vetrò et al., 2016)
Understandability	Percentagem de colunas com metadados, por Tabela	Tabela	Indica o valor da percentagem de colunas com metadados, por tabela	(Vetrò et al., 2016)

Na figura 3.7, pode ser verificado o exemplo do carregamento dos fluxos de carregamento das tabelas 'SA_Metric' e 'SA_TableOD'. Na figura 3.8, pode ser verificado o exemplo do carregamento da tabela 'SA_Measurement', em que são ilustrados os containers com o fluxo, quer por Tabela (*Datasets*), quer por Atributo (*Cell*), respeitando o nível de análise que consta na tabela 3.4.

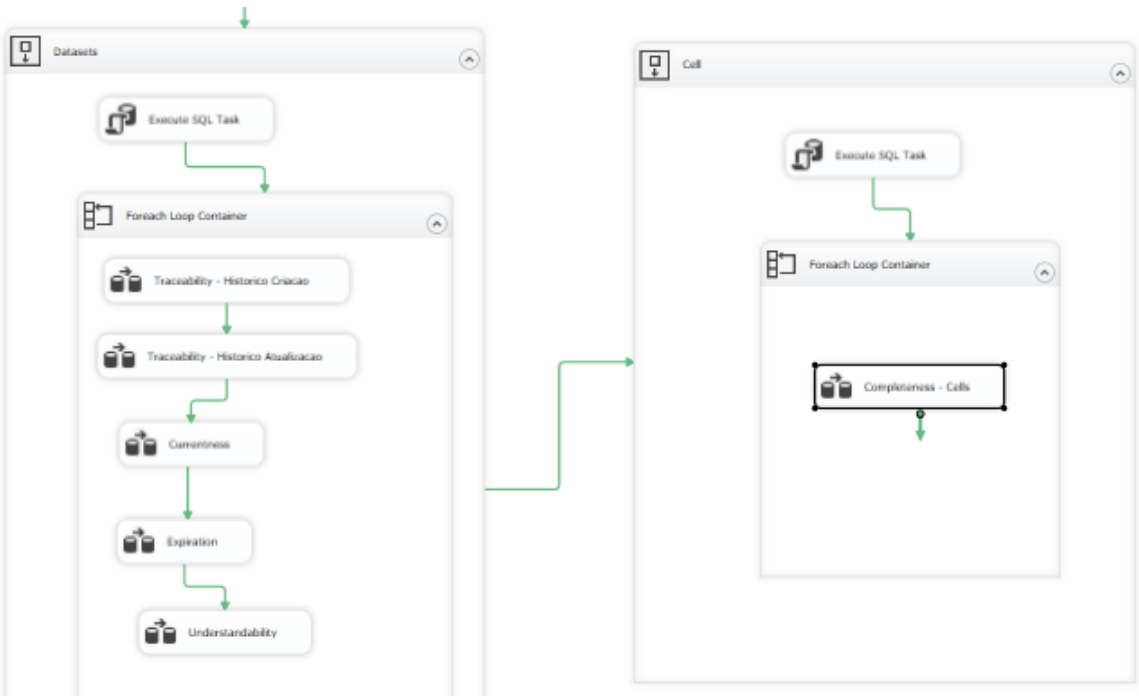

Figura 3.8 - Exemplo de carregamento por fluxos – SA *Load Fact Sources*

3.5.1.2. Control Flow – DW

O fluxo específico de carregamento dos dados na área de "Control Flow" em contexto de DW seguiu a seguinte ordem:

1. Efetuar a operação de carregamento da tabela de uma das dimensões, a tabela 'D_Metric'
2. Efetuar a operação de carregamento da tabela de uma das dimensões, a tabela 'D_TableOD'
3. Para ser evitada uma duplicação quando existe a necessidade de efetuar um reprocessamento, será efetuada uma operação de *delete* por 'Period_ID', procedendo-se posteriormente ao carregamento da tabela fatual, a tabela 'F_Measure'

Na figura 3.9, pode ser verificado o exemplo do carregamento dos fluxos de carregamento das tabelas 'D_Metric', 'D_TableOD' e da tabela 'F_Measure'.

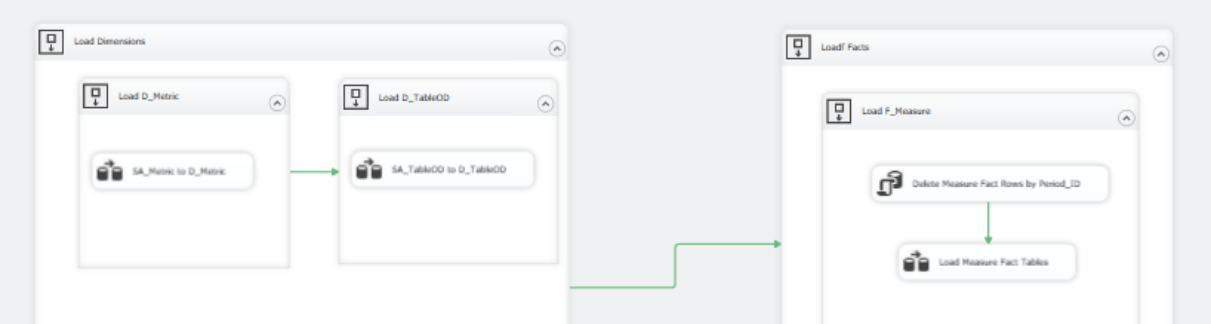

Figura 3.9 - Exemplo de carregamento em DW – DW *Load Dimensions* e *Load Facts*

3.5.2. Data Flow

Os processos de "Data Flow" seguem igualmente uma lógica idêntica de execução. Contudo, dadas as especificidades das fontes e das operações necessárias para executar, iremos apresentar as opções técnicas adotadas para resolver estes casos específicos. Em diversos casos foi tomada a opção de fazer transformações nos dados através de SQL nas *tasks* de *Sources*. Desta forma foi procurado obter uma *performance* superior em relação a efetuar a mesma transformação no SSIS, dado que a *performance* SQL é superior.

- **Inserção do 'Period_ID' no carregamento dos dados** - de forma a registar o 'Period_ID' do fluxo em execução nos dados a serem inseridos, foi utilizada a *task* "Derived Column", que obtém o valor da variável Period_ID, e a mapeia nos dados. (Figura 3.10).

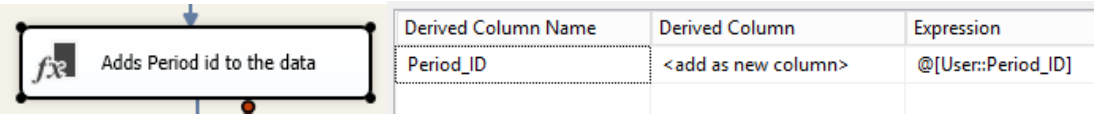

Figura 3.10 – Exemplo de utilização da *task* "Derived Column"

- Carregamentos incrementais das tabelas fatuais - de forma a elaborar um carregamento incremental da tabela fatual através do 'Period_ID' definido para o fluxo, criámos diversas variáveis do tipo *String*, que armazenam a query concatenada com a variável 'TableID' ou 'AttributeID', conforme o tipo de validação pretendido. Posteriormente na *task* de Source, é utilizada essa variável para definir a obtenção dos dados. (Figura 3.11)

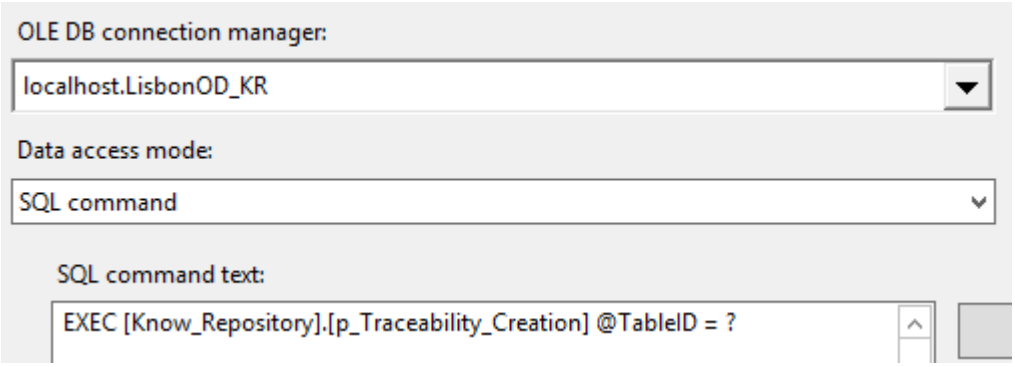

Figura 3.11 – Exemplo da Utilização de *Stored Procedure* com recurso a uma variável

4. RESULTADOS E DISCUSSÕES

Neste capítulo irão ser apresentados os resultados globais apurados do processo de avaliação da qualidade dos dados que irão ser colocados no Portal de Dados Abertos da cidade de Lisboa. Será também apresentado o *mockup* do *Dashboard* desenvolvido para mostrar os resultados dessa avaliação ao utilizador final.

4.1. DIMENSÕES/CARACTERÍSTICAS

Nesta secção irão se apresentados os resultados globais apurados do processo de avaliação da qualidade dos dados para cada uma das dimensões/características que foram estudadas neste Projeto.

4.1.1. *Traceability* – Histórico de Criação de um Conjunto de Dados

Conforme indicado na tabela 3.4, o apuramento da medição relativa à dimensão '*Traceability* – Histórico de criação da Tabela' indica a presença ou ausência de metadados associados ao processo de criação de um conjunto de dados.

Após a execução da métrica que apurou os resultados, foi verificado que, para a amostragem de conjuntos de dados (43) que foram analisados neste projeto (tabela 7.3), todos continham a presença de metadados associados ao seu processo de criação, como pode ser verificado na figura 4.1.

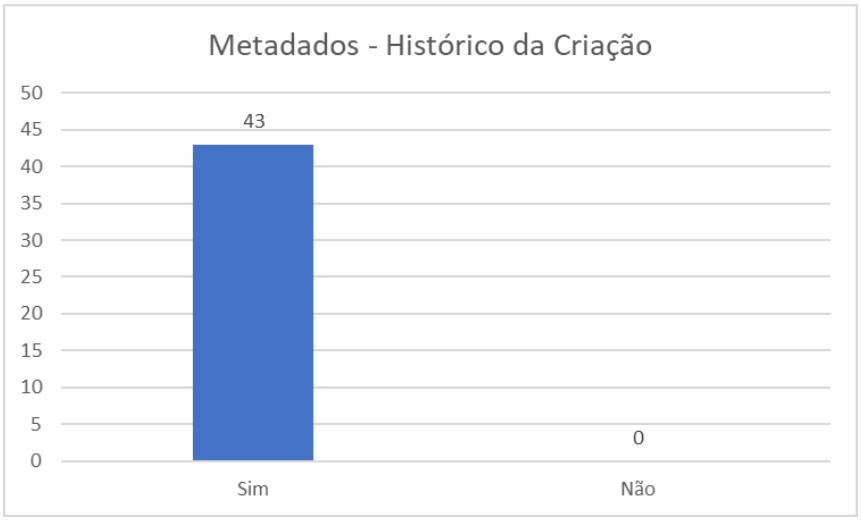

Figura 4.1 – Gráfico dos resultados da medição da Dimensão '*Traceability* – Histórico de criação de um Conjunto de Dados'

4.1.2. *Traceability* – Histórico de Atualizações de um Conjunto de Dados

Conforme indicado na tabela 3.4, o apuramento da medição relativa à dimensão '*Traceability* – Histórico de criação da Tabela' indica a presença ou ausência de metadados associados às atualizações efetuadas num conjunto de dados.

Após a execução da métrica que apurou os resultados, foi verificado que, para a amostragem de conjuntos de dados (43) que foram analisados neste projeto (tabela 7.3), todos continham a presença de metadados associados ao seu processo de atualizações, como pode ser verificado na figura 4.2.

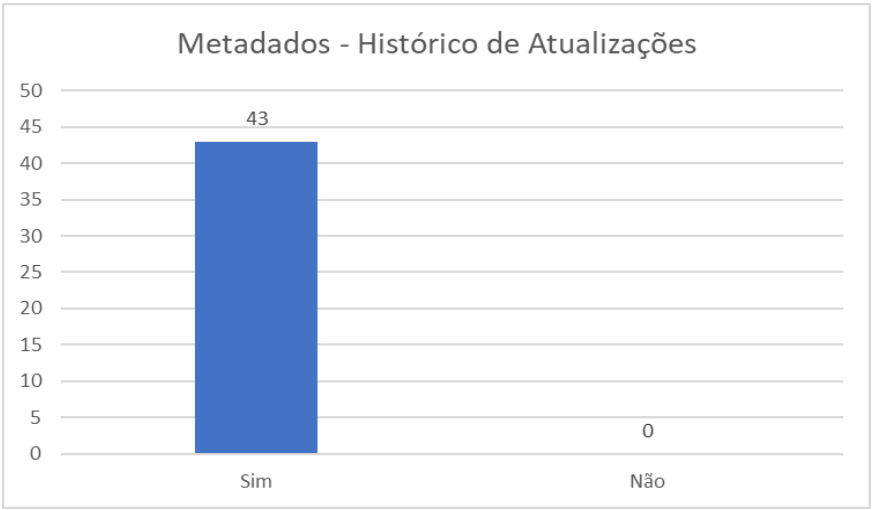

Figura 4.2 - Gráfico dos resultados da medição da Dimensão '*Traceability* – Histórico de atualizações de um Conjunto de Dados'

4.1.3. *Currentness* - Indicação de versão atualizada de um Conjunto de Dados

Conforme indicado na tabela 3.4, o apuramento da medição relativa à dimensão '*Currentness* – Indicação de versão atualizada de um Conjunto de Dados', que indica a atualidade da versão de um conjunto de dados.

Após a execução da métrica que apurou os resultados, foi verificado que, para a amostragem de conjuntos de dados (43) que foram analisados neste projeto (tabela 7.3), em 39 dos casos verificou-se que continham versões atualizadas, mas em 4 dos casos foi verificado que continham versões desatualizadas, como se pode verificar na figura 4.3. Na figura 4.4, é possível ainda verificar estes dados em termos de percentagem de Conjuntos de Dados com versões atualizadas e com versões desatualizadas.

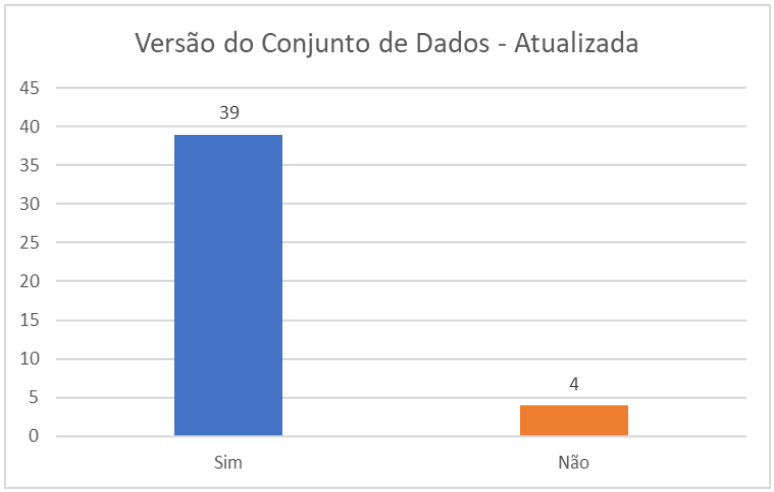

Figura 4.3 – Gráfico dos resultados da medição da Dimensão '*Currentness* – Indicação de versão atualizada de um Conjunto de Dados'

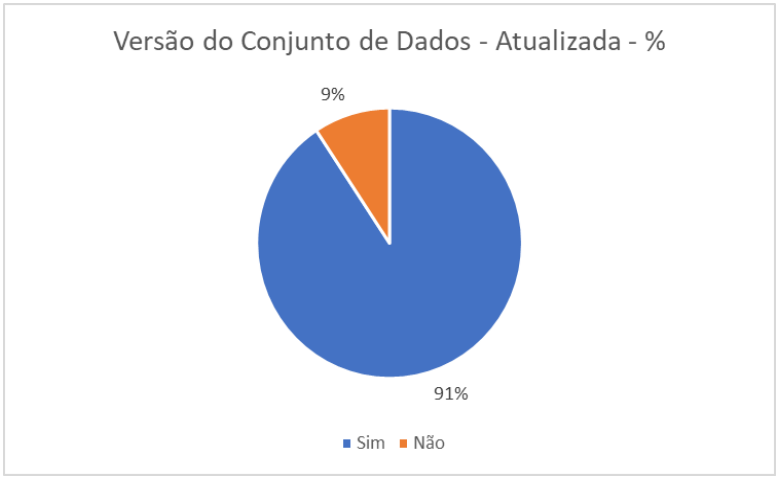

Figura 4.4 - Gráfico dos resultados da medição da Dimensão '*Currentness* – Indicação de versão atualizada de um Conjunto de Dados (%)'

4.1.4. *Expiration* - Indicação do atraso após a expiração da versão atual de um Conjunto de Dados

Conforme indicado na tabela 3.4, o apuramento da medição relativa à dimensão '*Expiration* – Indicação do atraso após a expiração da versão atual de um Conjunto de Dados', que indica a relação entre o atraso na publicação de um conjunto de dados após a data de expiração da sua versão anterior e o período de tempo referido pelo conjunto de dados, em meses.

Após a execução da métrica que apurou os resultados, foi verificado que, para a amostragem de conjuntos de dados (43) que foram analisados neste projeto (tabela 7.3), em 39 dos casos verificou-se que continham versões atualizadas, mas em 4 dos casos foi verificado que continham versões desatualizadas. Em termos de atraso após a expiração da versão atual, desses 4 casos, 3 conjuntos de

dados apresentaram uma versão desatualizada de 3 meses e 1 conjunto de dados apresentou uma versão desatualizada de 21 meses, conforme se pode verificar na figura 4.5. Na figura 4.6, é possível ainda verificar estes dados em termos de percentagem.

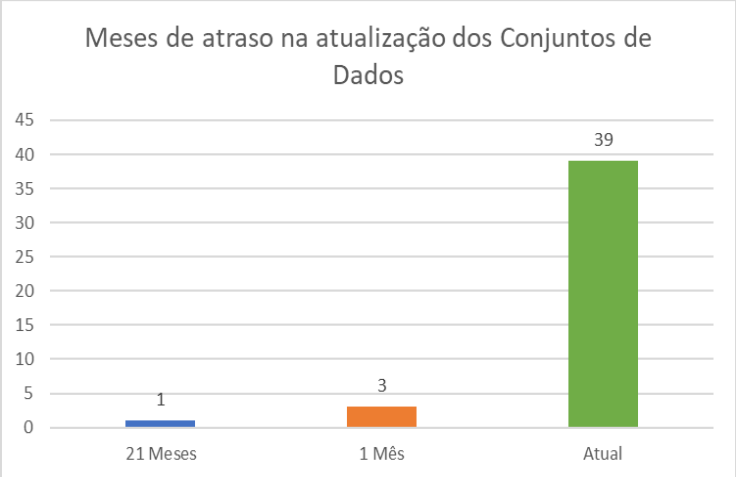

Figura 4.5 – Gráfico dos resultados da medição da Dimensão '*Expiration* – Indicação do atraso após a expiração da versão atual de um Conjunto de Dados'

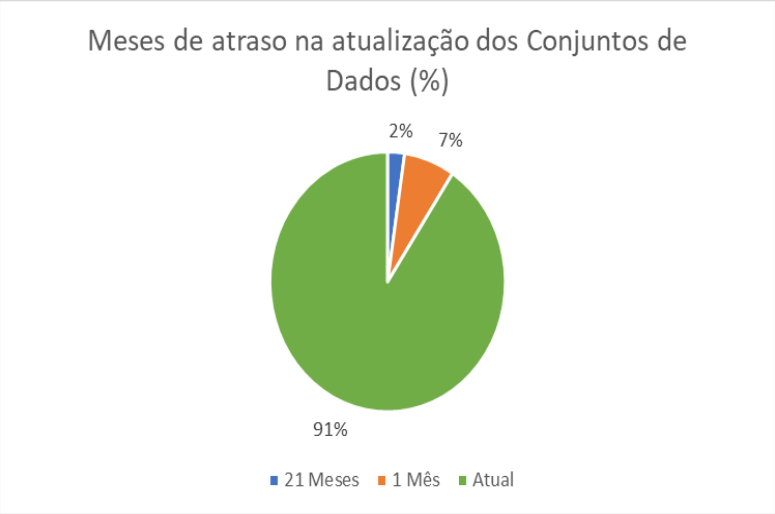

Figura 4.6 - Gráfico dos resultados da medição da Dimensão '*Expiration* – Indicação do atraso após a expiração da versão atual de um Conjunto de Dados (%)'

4.1.5. *Completeness* – Indicação da percentagem de células que não estão vazias em cada coluna dos Conjuntos de Dados

Conforme indicado na tabela 3.4, o apuramento da medição relativa à dimensão '*Completeness* – Indicação da percentagem de células que estão preenchidas em cada coluna dos Conjuntos de Dados', que indica a percentagem de células que não têm qualquer valor, em cada coluna dos conjuntos de dados que foram analisados.

Após a execução da métrica que apurou os resultados, foi verificado que, para a amostragem de conjuntos de dados (43) que foram analisados neste projeto (tabela 7.3) e para um total de 522 colunas ou atributos, em 426 colunas nenhuma das suas células estava vazia. No ponto oposto, verifica-se que 43 colunas não apresentavam qualquer valor preenchido. No gráfico da figura 4.7, pode-se verificar a distribuição da percentagem de células que estão preenchidas em cada coluna dos Conjuntos de Dados.

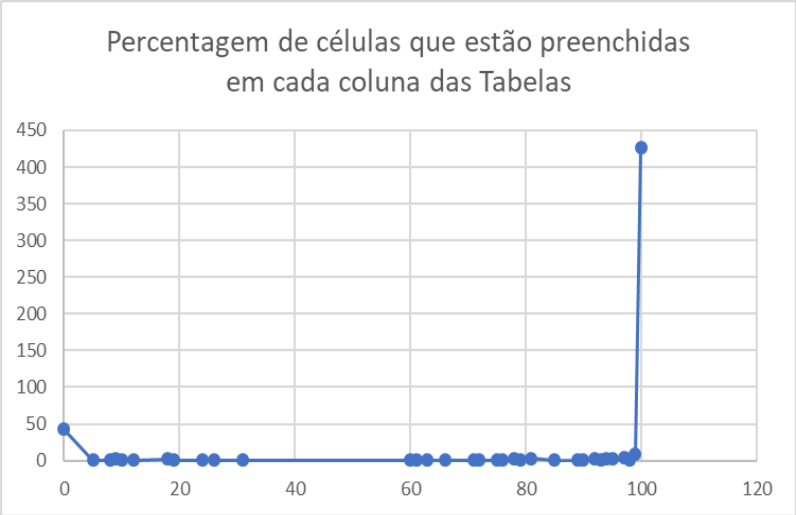

Figura 4.7 - Gráfico dos resultados da medição da Dimensão '*Completeness* – Indicação da percentagem de células que estão preenchidas em cada coluna dos Conjuntos de Dados'

4.1.6. *Understandability* - Indicação do valor da percentagem de colunas com metadados associados, por Conjunto de Dados

Conforme indicado na tabela 3.4, o apuramento da medição relativa à dimensão '*Understandability* – Indicação do valor da percentagem de colunas com metadados associados, por Conjuntos de Dados', que indica a percentagem de colunas que têm metadados associados, nos conjuntos de dados que foram analisados. Para cada conjunto de dados foram avaliadas 8 colunas que contêm metadados associados.

Após a execução da métrica que apurou os resultados, foi verificado que, para a amostragem de conjuntos de dados (43) que foram analisados neste projeto (tabela 7.3), em 27 casos foi verificado que todas as colunas que contêm metadados associados estão preenchidas. Na figura 4.8, pode-se

verificar a distribuição da percentagem de colunas com metadados associados, por conjunto de dados.

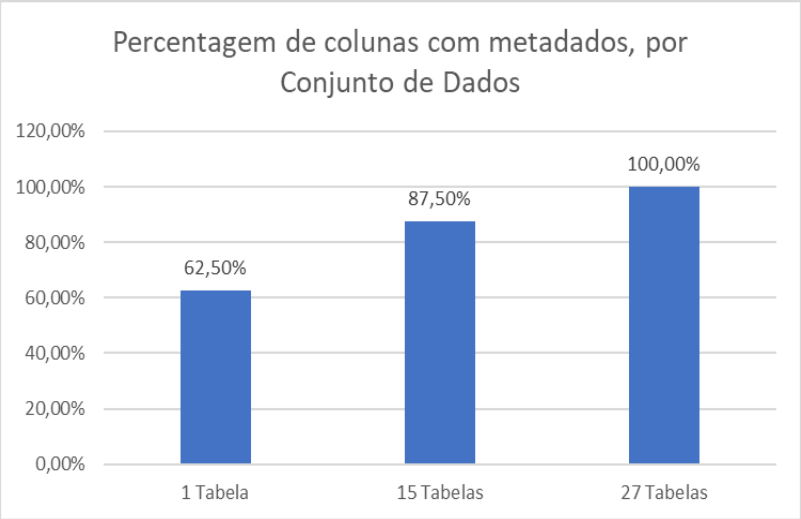

Figura 4.8 - Gráfico dos resultados da medição da Dimensão '*Understandability* – Indicação do valor da percentagem de colunas com metadados associados, por Conjuntos de Dados'

4.2. DASHBOARD – MOCKUP

Após o apuramento dos resultados globais apurados do processo de avaliação da qualidade dos dados que irão ser colocados no Portal de Dados Abertos da cidade de Lisboa, foi desenvolvido um *mockup* de um *Dashboard* (figura 4.9) na ferramenta 'Power BI', que permita ao utilizador final uma visualização rápida e bastante compreensível dos resultados dessa avaliação. O desenho deste *mockup* teve como base o *wireframe* descrito na figura 2.4., do capítulo 2.4.1 deste documento.

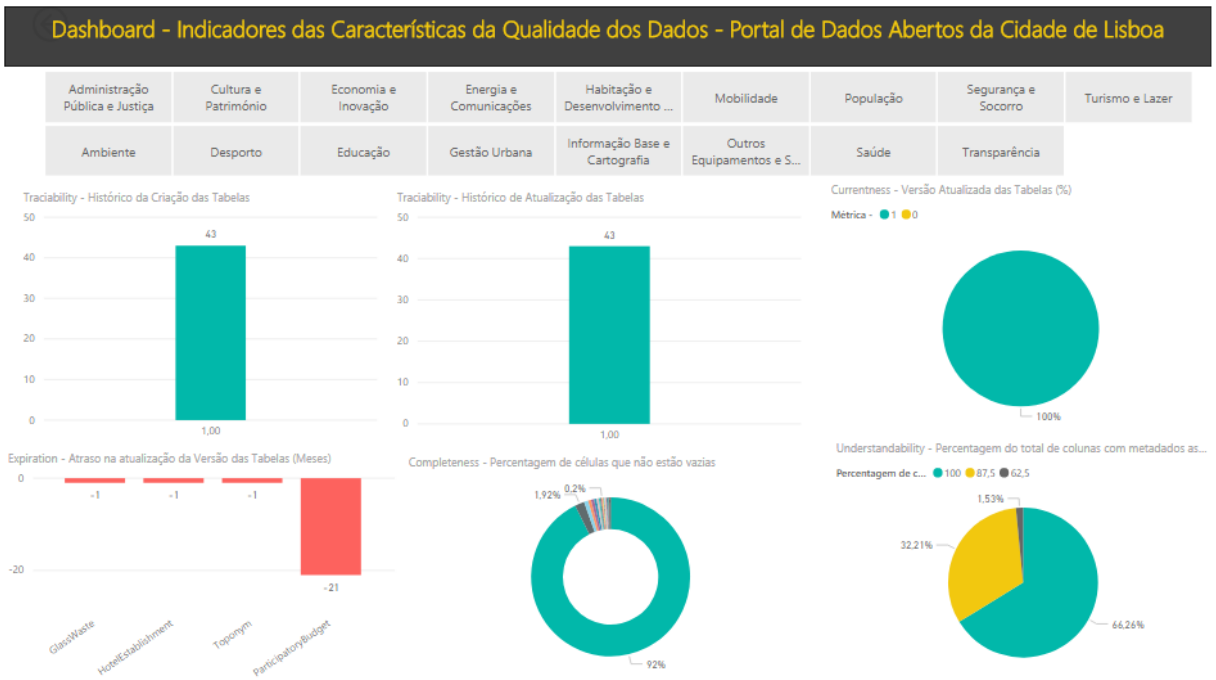

Figura 4.9 – *Mockup* do *Dashboard* com os indicadores da Qualidade dos Dados que foram avaliados na *framework* de avaliação da qualidade dos dados

Na figura 4.9, pode-se verificar que neste *mockup* surgem gráficos com resultados de todas as dimensões/características que foram analisadas neste projeto. O utilizador também tem a possibilidade de selecionar os resultados por grupo de dados, através do *slicer* que consta na parte superior do *Dashboard*. Esses grupos de dados são os 18 grupos de dados que constam no Portal de Dados Abertos da cidade de Lisboa.

4.3. AVALIAÇÃO DOS RESULTADOS

Após o apuramento dos indicadores descritos nos pontos acima, para a amostra de conjuntos de dados que foram analisados (43) e para as dimensões/características que foram estudadas, pode-se verificar que, de uma forma geral, os dados que são colocados no Portal de Dados Abertos da Cidade de Lisboa têm boa qualidade.

Ao nível dos indicadores relacionados com a presença de metadados, quer sejam relacionados com as tabelas (*Traceability*), quer sejam relacionados com as colunas que constam nessas mesmas tabelas (*Understandability*), pode-se verificar que a falta de metadados é praticamente residual, em relação ao total apurado.

Relativamente à atualidade das versões dos conjuntos de dados que foram analisados, pode-se verificar que em cerca de 91% da amostra as versões encontram-se atualizadas. Assim sendo, existirá alguma margem para a atualização das versões que não se encontram atualizadas, de modo a promover a possibilidade da sua reutilização pelos utilizadores.

O apuramento do valor do número de células que não estão vazias em cada coluna dos conjuntos de dados verificados, permite verificar que, cerca de 82% das colunas verificadas continham todas as células preenchidas. Cerca de 8% das colunas não continham qualquer valor nas suas células, mas este facto tem de ser validado conforme o contexto (requisito de qualidade) em que esse atributo se encontra no conjunto de dados, de forma a verificar se será uma falha ou não.

5. CONCLUSÕES

O desenvolvimento deste Projeto teve como objetivo principal o desenvolvimento de uma *framework* para a avaliação e o tratamento da qualidade dos dados que são colocados no portal de Dados Abertos da cidade de Lisboa, de modo a promover uma maior reutilização por parte dos cidadãos para a criação de novos produtos e serviços.

Seguindo os passos que foram definidos na metodologia DSRM, foi desenvolvido um artefacto de IT que permite analisar algumas dimensões/características de qualidade dos dados, de modo a atingir o objetivo estabelecido para este Projeto.

Na revisão de literatura que foi levada a cabo é possível verificar que há evidências de um crescimento da influência da presença de Dados Abertos com qualidade, na evolução da implementação de iniciativas privadas e públicas, que possam promover a criação de novos produtos e serviços para benefício da comunidade.

O estudo levado a cabo sobre os dados que são colocados no Portal de Dados Abertos da cidade de Lisboa permitiu verificar que existe um grande volume de dados que são disponibilizados, em muitas áreas de atividade, pelo que existe uma grande motivação para que a qualidade dos dados que são disponibilizados seja apurada e, posteriormente, melhorada.

Os diversos formatos dos dados que são colocados no Portal de Dados Abertos levou a um trabalho exaustivo de preparação dos mesmos, e foram todos colocados em bases de dados em SQL Server para poderem ser avaliados na *framework* que foi implementada.

Foi efetuado um estudo das diversas metodologias de avaliação da qualidade dos dados que se encontram disponíveis na literatura e das suas métricas de avaliação. Foram verificadas e estudadas diversas métricas, o seu contexto de aplicação e as suas características e foi selecionado um conjunto das mesmas, para serem implementadas neste Projeto.

Após a implementação do cálculo das métricas, foi desenvolvido ainda um *mockup* de um *Dashboard*, para implementação de um *Dashboard* para monitorização da qualidade dos dados que são avaliados através desta *framework*.

A implementação do cálculo das métricas foi efetuada com recurso a ferramentas Microsoft, em que na elaboração do DW foi utilizado o '*SQL Server Management Studio 2016*' (SSMS). Os processos de ETL foram definidos em SQL Server Integration Services (SSIS) através da ferramenta de desenvolvimento '*SQL Server Data Tools 2015*' e o desenvolvimento do *mockup* do *Dashboard* foi efetuado com recurso à ferramenta '*Power BI*'.

Após o apuramento do valor das métricas de avaliação da qualidade de dados sobre o conjunto de dados utilizados, pode-se concluir que, pelo menos mediante estas métricas, a qualidade dos dados que são colocados no Portal de Dados Abertos da Cidade de Lisboa têm boa qualidade.

Assim sendo, pode-se concluir que os objetivos principais e específicos que foram determinados no início deste projeto foram atingidos.

5.1. LIMITAÇÕES DO PROJETO

As maiores limitações encontradas no desenvolvimento deste Projeto foram as seguintes:

- Em muitas das métricas que foram estudadas, existiram bastantes que só poderiam ser calculadas conforme o contexto (requisito de qualidade) em que os dados iriam ser utilizados
- Para o mesmo conjunto de dados, existem métricas que iriam indicar uma boa qualidade desse conjunto de dados num determinado contexto, mas se o contexto for alterado, o cálculo da métrica também iria ser alterado e nesse caso já poderia indicar um resultado diferente

5.2. RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Existem diversos pontos em que o trabalho desenvolvido neste projeto pode ser melhorado, nomeadamente:

- Proceder ao desenvolvimento e aplicação de outras métricas de avaliação da qualidade dos dados que não foram implementadas neste projeto
- Dentro dessas métricas, procurar saber com o utilizador final dos conjuntos de dados, qual o contexto em que os dados irão ser utilizados, para cada caso
- Proceder à avaliação de mais conjuntos de dados que são colocados no Portal de Dados Abertos da cidade de Lisboa, além da amostra que foi avaliada neste projeto, para o apuramento de um resultado geral sobre a qualidade dos dados
- Tendo como base o *mockup* do *Dashboard* que foi desenvolvido, deve ser desenvolvido um *Dashboard* mais completo, com mais detalhe e com mais métricas e indicadores, para que o utilizador final consiga verificar quais os dados que necessitam de intervenção para que a sua qualidade seja melhorada, antes de serem disponibilizados no Portal de Dados Abertos
- Construção de cubos OLAP sobre as tabelas factuais e dimensões implementadas nesta fase de desenvolvimento
- Construção de um *layer* de *reporting* assente nos cubos, de forma a disponibilizar os dados de uma forma simples e clara para os utilizadores do negócio

6. BIBLIOGRAFIA

Aguilera, U., Peña, O., Belmonte, O., & López-de-Ipiña, D. (2016). Citizen-centric data services for smarter cities. *Future Generation Computer Systems*, *76*, 234–247. https://doi.org/10.1016/j.future.2016.10.031

Albino, V., Berardi, U., & Dangelico, R. M. (2015). Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, *22*(1), 1–19. https://doi.org/10.1080/10630732.2014.942092

AMA - Agência para a Modernização Administrativa, I. P. (n.d.). Portal de dados abertos da Administração Pública. Retrieved November 12, 2018, from https://dados.gov.pt/pt/

Anandarajan, M., & Jones, D. (2017). Why balance is the key to dashboards. *CIO.*, p27–28. 2p. Retrieved from https://eds.a.ebscohost.com/eds/detail/detail?vid=0&sid=e3f41ee5-31e5-4f2b-8310-d67c281203ed%40sessionmgr4007&bdata=JkF1dGhUeXBlPWlwLGNvb2tpZSxzaGliLHVpZCZsYW5nPXB0LWJyJnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ%3D%3D#AN=123549965&db=bth

Angelidou, M. (2014). Smart city policies: A spatial approach. *Cities*, *41*, S3–S11. https://doi.org/10.1016/j.cities.2014.06.007

Ardakan, M. A., & Mohajeri, K. (2009). Applying design research method to IT performance management: Forming a new solution. *Journal of Applied Sciences*, *9*(7), 1227–1237. https://doi.org/10.3923/jas.2009.1227.1237

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, *32*(4), 399–418. https://doi.org/10.1016/j.giq.2015.07.006

Bakici, T., Almirall, E., & Wareham, J. (2013). A Smart City Initiative: The Case of Barcelona. *Journal of the Knowledge Economy*, *4*(2), 135–148. https://doi.org/10.1007/s13132-012-0084-9

Batini, C., Barone, D., Mastrella, M., Maurino, A., & Ruffini, C. (2007). A FRAMEWORK AND A METHODOLOGY FOR DATA QUALITY ASSESSMENT AND MONITORING. *Conference: Proceedings of the 12th International Conference on Information Quality, MIT, Cambridge, MA, USA, November 9-11, 2007*. Retrieved from https://pdfs.semanticscholar.org/066b/124ced2a6a5330bed8da00584352f5c19cd9.pdf

Batini, C., Cabitza, F., Cappiello, C., & Francalanci, C. (2006). A comprehensive data quality methodology for web and structured data. In *2006 1st International Conference on Digital Information Management, ICDIM* (pp. 448–456). IEEE. https://doi.org/10.1109/ICDIM.2007.369236

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, *41*(3), 1–52. https://doi.org/10.1145/1541880.1541883

Bellavista, P., Tortonesi, M., Lanzon, S., Riberto, G., Stefanelli, C., & Tortonesi, M. (2017). A middleware solution for wireless iot applications in sparse smart cities. *Sensors (Switzerland)*, *17*(11), 2525. https://doi.org/10.3390/s17112525

Bicevskis, J., Bicevska, Z., & Karnitis, G. (2017). Executable Data Quality Models. *Procedia Computer Science*, *104*, 138–145. https://doi.org/10.1016/j.procs.2017.01.087

Bizer, C., & Cyganiak, R. (2009). Quality-driven information filtering using the WIQA policy framework. *Web Semantics*, *7*(1), 1–10. https://doi.org/10.1016/j.websem.2008.02.005

Borovina Josko, J. M., & Ferreira, J. E. (2017). Visualization properties for data quality visual assessment: An exploratory case study. *Information Visualization*, *16*(2), 93–112. https://doi.org/10.1177/1473871616629516

Bovee, M., Srivastava, R. P., & Mak, B. (2003). A Conceptual Framework and Belief- Function Approach to Assessing Overall Information Quality. In *Proceedings of the Sixth International Conference on Information Quality* (Vol. 18, pp. 311–328). https://doi.org/10.1002/int.10074

Calero, C., Caro, A., & Piattini, M. (2008). An applicable data quality model for web portal data consumers. *World Wide Web*, *11*(4), 465–484. https://doi.org/10.1007/s11280-008-0048-y

Câmara Municipal de Lisboa. (n.d.). Portal Dados Abertos da Cidade de Lisboa. Retrieved July 29, 2018, from http://dados.cm-lisboa.pt/

Cappiello, C., Francalanci, C., & Pernici, B. (2004). Data quality assessment from the user's perspective. *Proceedings of the 2004 International Workshop on Information Quality in Information Systems*, 68–73. https://doi.org/10.1145/1012453.1012465

Caragliu, A., Del Bo, C., & Nijkamp, P. (2011). Smart Cities in Europe. *Journal of Urban TechnologyOnline) Journal Smart Cities in Europe Journal of Urban Technology*, *18*(2), 1063–732. https://doi.org/10.1080/10630732.2011.601117

Carlo, B., Daniele, B., Federico, C., & Simone, G. (2011). A DATA QUALITY METHODOLOGY FOR HETEROGENEOUS DATA. *International Journal of Database Management Systems (IJDMS)*, *3*(1). https://doi.org/10.5121/ijdms.2011.3105

Carrara, W., Oudkerk, F., Steenbergen, E. van, & Tinholt, D. (2016). Open Data Goldbook for Data Managers and Data Holders. Retrieved from https://www.europeandataportal.eu/sites/default/files/goldbook.pdf

Carrara, W., Radu, C., & Vollers, H. (2017). Open Data Maturity in Europe 2017. *European Data Portal*. Retrieved from https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n3_2017.pdf

Carrara, W., San Chan, W., Fischer, S., & van Steenbergen, E. (2015). *Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources*. https://doi.org/10.2759/328101

Carretero, A. G., Gualo, F., Caballero, I., & Piattini, M. (2017). MAMD 2.0: Environment for data quality processes implantation based on ISO 8000-6X and ISO/IEC 33000. *Computer Standards and Interfaces*, *54*, 139–151. https://doi.org/10.1016/j.csi.2016.11.008

CRETU, L.-G. (2012). Smart Cities Design using Event-driven Paradigm and Semantic Web. *Informatica Economica*, *16*(4), 57–67. Retrieved from https://pdfs.semanticscholar.org/2f44/6982f70863778981977ae3f287a5cd8adbc5.pdf

De Amicis, F., Barone, D., & Batini, C. (2006). AN ANALYTICAL FRAMEWORK TO ANALYZE DEPENDENCIES AMONG DATA QUALITY DIMENSIONS. *Proceedings of the 11th International Conference on Information Quality (ICIQ)*. Retrieved from http://mitiq.mit.edu/ICIQ/Documents/IQ Conference 2006/Papers/An Analytical Framework to Analyze Dependencies Among Data Quality Dimensions.pdf

de Castro Neto, M., Rego, J. S., Neves, F. T., & Cartaxo, T. M. (2017). Smart open cities: Portuguese municipalities open data policies evaluation. In *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1–6). IEEE. https://doi.org/10.23919/CISTI.2017.7975912

Ehsan, H., Sharaf, M. A., & Chrysanthis, P. K. (2016). MuVE: Efficient Multi-Objective View Recommendation for Visual Data Exploration. In *2016 IEEE 32nd International Conference on Data Engineering, ICDE 2016* (pp. 731–742). IEEE. https://doi.org/10.1109/ICDE.2016.7498285

English, L. P. (1999). *Improving data warehouse and business information quality : methods for reducing costs and increasing profits*. Wiley. Retrieved from https://dl.acm.org/citation.cfm?id=299503

Eppler, M. J., & Helfert, M. (2004). A classification and analysis of data quality costs. *Proceedings of the Ninth International Conference on Informationen (ICIQ-04)*, 311–325. Retrieved from https://pdfs.semanticscholar.org/02ef/a0fb30d72a587d5531ddeb360f71e02c5704.pdf

Even, A., & Shankaranarayanan, G. (2007). Utility-Driven Assessment of Data Quality. *The DATA BASE for Advances in Information Systems*, *38*(2), 75–93. https://doi.org/10.1145/1240616.1240623

Falorsi, P.D.; Pallara, S.; Pavone, A.; Alessandroni, A.; Massella, E.; and Scannapieco, M. (2003). Improving the quality of toponymic data in the italian public administration. *Proceedings of the ICDTWorkshop on Data Quality in Cooperative Information Systems (DQCIS)*.

Few, S. (2007). Data Visualization: Past, Present, and Future. *IBM Cognos Innovation Center for Performance Management*, 3–11. Retrieved from https://www.perceptualedge.com/articles/Whitepapers/Data_Visualization.pdf

Geerts, G. L. (2011). A design science research methodology and its application to accounting information systems research. *International Journal of Accounting Information Systems*, *12*(2), 142–151. https://doi.org/10.1016/j.accinf.2011.02.004

Grossi, G., & Pianezzi, D. (2017). Smart cities: Utopia or neoliberal ideology? *Cities*, *69*, 79–85. https://doi.org/10.1016/j.cities.2017.07.012

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., … Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management*, *36*(5), 748–758. https://doi.org/10.1016/j.ijinfomgt.2016.05.002

Heinrich, B., Kaiser, M., & Klier, M. (2011). How To Measure Data Quality ? In *Proceedings of the 16th International Conference on Information Quality (ICIQ-2011)* (pp. 1–15). Retrieved from https://epub.uni-regensburg.de/23633/1/heinrich.pdf

Heinrich, B., Klier, M., & Kaiser, M. (2009). A Procedure to Develop Metrics for Currency and its Application in CRM. *Journal of Data and Information Quality*, *1*(1), 1–28. https://doi.org/10.1145/1515693.1515697

Hevner, A., & Chatterjee, S. (2010). Introduction to Design Science Research, (July), 1–8. https://doi.org/10.1007/978-1-4419-5653-8_1

Hossain, M. A., Dwivedi, Y. K., & Rana, N. P. (2016). State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 14–40. https://doi.org/10.1080/10919392.2015.1124007

ISO/IEC 25012. (2008). ISO/IEC 25012:2008 - Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Data quality model. Retrieved February 12, 2018,

from https://www.iso.org/standard/35736.html

ISO 9001. (2008). Sistemas de da qualidade. Requisitos (ISO 9001:2008), *406*, 2008–11. Retrieved from https://www.iso.org/standard/46486.html

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. https://doi.org/10.1080/10580530.2012.716740

Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, *70*, 338–345. https://doi.org/10.1016/j.jbusres.2016.08.007

Jetzek, T. (2016). Managing complexity across multiple dimensions of liquid open data: The case of the Danish Basic Data Program. *Government Information Quarterly*, *33*(1), 89–104. https://doi.org/10.1016/j.giq.2015.11.003

Jeusfeld, M. A., Quix, C., & Jarke, M. (1998). Design and Analysis of Quality Information for Data Warehouses. In *Proceedings of the 17th International Conference on Conceptual Modeling* (pp. 349–362). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-49524-6_28

Juice. (2009). *A Guide to Creating Dashboards People Love to Use*. *Juice Analytics*. Retrieved from http://www/juiceanalytics.com/poster/

Karkouch, A., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Data quality in internet of things: A state-of-the-art survey. *Journal of Network and Computer Applications*, *73*, 57–81. https://doi.org/10.1016/j.jnca.2016.08.002

Kleindienst, D. (2017). The data quality improvement plan: deciding on choice and sequence of data quality improvements. *Electronic Markets - The International Journal on Networked Business*, *27*(4), 1–12. https://doi.org/10.1007/s12525-017-0245-6

Kreitzberg, C. B. (2004). *Strategic Business Leaders Reducing Outsourcing Risk Through Visual Communication*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.86.1906&rep=rep1&type=pdf

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: A methodology for information quality assessment. *Information and Management*, *40*(2), 133–146. https://doi.org/10.1016/S0378-7206(02)00043-5

Long, J., & Seko, C. (2002). A New Method for Database Data Quality Evaluation at the Canadian Institute for Health Information (CIHI). *Proceedings of 7th International Conference on Information Quality*. Retrieved from www.statcan.ca

Loshin, D. (2001). *Enterprise Knowledge Management. The Data Quality Approach*. Morgan Kaufmann.

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly*, *32*(3), 323–332. https://doi.org/10.1016/j.giq.2015.05.006

Máchová, R., & Lněnička, M. (2017). Evaluating the quality of open data portals on the national level. *Journal of Theoretical and Applied Electronic Commerce Research*, *12*(1), 21–41. https://doi.org/10.4067/S0718-18762017000100003

Marco, A. De, Mangano, G., & Zenezini, G. (2015). Digital Dashboards for Smart City Governance: A Case Project to Develop an Urban Safety Indicator Model. *Journal of Computer and Communications*, *03*(05), 144–152. https://doi.org/10.4236/jcc.2015.35018

Marsal-Llacuna, M. L., Colomer-Llinàs, J., & Meléndez-Frigola, J. (2014). Lessons in urban monitoring taken from sustainable and livable cities to better address the Smart Cities initiative. *Technological Forecasting and Social Change*, *90*(PB), 611–622. https://doi.org/10.1016/j.techfore.2014.01.012

Merino, J., Caballero, I., Rivas, B., Serrano, M., & Piattini, M. (2016). A Data Quality in Use model for Big Data. *Future Generation Computer Systems*, *63*, 123–130. https://doi.org/10.1016/j.future.2015.11.024

Moraga, C., Moraga, M. Á., Calero, C., & Caro, Á. (2009). SQuaRE-Aligned Data Quality Model for Web Portals. *2009 Ninth International Conference on Quality Software*, 117–122. https://doi.org/10.1109/QSIC.2009.23

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pereira, G. V., Macadar, M. A., Luciano, E. M., & Testa, M. G. (2017). Delivering public value through open government data initiatives in a Smart City context. *Information Systems Frontiers*, *19*(2), 213–229. https://doi.org/10.1007/s10796-016-9673-7

Pipino, L. L., Lee, Y. W., Wang, R. Y., Lowell Yang Lee, M. W., & Yang, R. Y. (2002). Data Quality Assessment. *Communications of the ACM*, *45*(4), 211. https://doi.org/10.1145/505248.506010

Rathore, M. M., Ahmad, A., Paul, A., & Rho, S. (2016). Urban planning and building smart cities based on the Internet of Things using Big Data analytics. *Computer Networks*, *101*, 63–80. https://doi.org/10.1016/j.comnet.2015.12.023

Rivero, J. M., Grigera, J., Rossi, G., Robles Luna, E., Montero, F., & Gaedke, M. (2014). Mockup-Driven Development: Providing agile support for Model-Driven Web Engineering. *Information and Software Technology*, *56*(6), 670–687. https://doi.org/10.1016/J.INFSOF.2014.01.011

Ruijer, E., Grimmelikhuijsen, S., & Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. *Government Information Quarterly*, *34*(1), 45–52. https://doi.org/10.1016/j.giq.2017.01.001

Rula, A., & Zaveri, A. (2014). Methodology for Assessment of Linked Data Quality, 1–4. Retrieved from https://pdfs.semanticscholar.org/1a6d/d2d67acbbecca1fe635b4bac099b1943818e.pdf

Sadiq, S., & Indulska, M. (2017). Open data: Quality over quantity. *International Journal of Information Management*, *37*(3), 150–154. https://doi.org/10.1016/j.ijinfomgt.2017.01.003

Scannapieco, M., Virgillito, A., Marchetti, C., Mecella, M., & Baldoni, R. (2004). The DaQuinCIS architecture: A platform for exchanging and improving data quality in cooperative information systems. *Information Systems*, *29*(7), 551–582. https://doi.org/10.1016/j.is.2003.12.004

Sieber, R. E., & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, *32*(3), 308–315. https://doi.org/10.1016/j.giq.2015.05.003

Sta, H. Ben. (2017). Quality and the efficiency of data in “Smart-Cities.” *Future Generation Computer*

Systems, *74*, 409–416. https://doi.org/10.1016/j.future.2016.12.021

Su, Z., & Jin, Z. (2004). A Methodology for Information Quality Assessment in the Designing and Manufacturing Processes of Mechanical Products. In *Proceedings of the Ninth International Conference on Information Quality (ICIQ-04)* (pp. 190–220). https://doi.org/10.4018/978-1-59904-024-0.ch009

Thorsby, J., Stowers, G. N. L., Wolslegel, K., & Tumbuan, E. (2017). Understanding the content and features of open data portals in American cities. *Government Information Quarterly*, *34*(1), 53–61. https://doi.org/10.1016/j.giq.2016.07.001

Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. *OECD Working Papers on Public Governance*, (22). https://doi.org/10.1787/5k46bj4f03s7-en

Verhulst, S., Young, A., Mcmurren, J., Noveck, B. S., Rogawksi, C., & Sangokoya, D. (2016). Open Data impact, when demand and supplymeet. Key finding of the open data Impact case studies. Retrieved from www.odimpact.org

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, *33*(2), 325–337. https://doi.org/10.1016/j.giq.2016.02.001

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86–95. https://doi.org/10.1145/240455.240479

Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM*, *41*(2), 58–65. https://doi.org/10.1145/269012.269022

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

www.datawarehouse4u.info. (n.d.). Data Warehouse Schema Architecture - star schema. Retrieved August 12, 2018, from http://datawarehouse4u.info/Data-warehouse-schema-architecture-star-schema.html

Yeganeh, N. K., Sadiq, S., & Sharaf, M. A. (2014, December 1). A framework for data quality aware query systems. *Information Systems*. Pergamon. https://doi.org/10.1016/j.is.2014.05.005

Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, *13*(1), 41–59. https://doi.org/10.1016/j.accinf.2011.08.002

Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., & Auer, S. (2012). Quality Assessment Methodologies for Linked Open Data: A Systematic Literature Review and Conceptual Framework. *Semantic Web – Interoperability, Usability, Applicability*, *1*, 33. https://doi.org/10.3233/SW-150175

Zotano, M. A. G., & Bersini, H. (2017). A Data-driven Approach to Assess the Potential of Smart Cities: The Case of Open Data for Brussels Capital Region. In *Energy Procedia* (Vol. 111, pp. 750–758). https://doi.org/10.1016/j.egypro.2017.03.237

Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, *19*(1–2), 17–33. https://doi.org/10.3233/IP-140329

7. APENDICE

Tabela 7.1 – Tabela com a descrição das Dimensões/Características encontradas na revisão de literatura

Dimensões / Características	Descrição	Referências
Accessibility	O grau em que um portal Web fornece uma navegação simples e acessível, de modo a que o utilizador consiga obter os dados de uma forma rápida e fácil.	(Calero et al., 2008; Moraga et al., 2009)
Accuracy	O grau que indica a medida em que os dados estão livres de erros, estão corretos e são confiáveis.	(Calero et al., 2008; Moraga et al., 2009)
Adaptability	O grau que indica a capacidade de adaptação. É uma característica da metodologia CIHI.	(Long & Seko, 2002)
Agreement of Usage	Acordo de Utilização dos Dados.	(Loshin, 2001)
Amount of Data	O grau que indica a medida em que a quantidade ou volume de dados entregue pelo portal Web é apropriado para a tarefa em questão.	(Calero et al., 2008; Moraga et al., 2009; Pipino et al., 2002)
Applicability	O grau que indica a medida em que os dados são específicos, úteis e fáceis de aplicar para os utilizadores alvo.	(Calero et al., 2008; Moraga et al., 2009)
Appropriateness	O grau que indica a quantidade de informação adequada a uma determinada utilização.	(Su & Jin, 2004; Wang, 1998)
Arrangement	O grau que indica a disposição ou ordenação dos dados	(Su & Jin, 2004)
Attractiveness	O grau que indica a medida em que os dados que constam no portal são atraentes para seus visitantes	(Calero et al., 2008; Moraga et al., 2009)
Attribute granularity	O grau que indica a medida de granularidade dos atributos dos *schemas* onde estão guardados os dados.	(Loshin, 2001)
Availability	O grau que indica a medida em que os dados estão disponíveis através do portal Web e possuem atributos que os permitem ser utilizados por utilizadores e aplicações autorizadas.	(Calero et al., 2008; Moraga et al., 2009)
Believability	O grau que indica a medida em que os dados e a sua	(Calero et al., 2008)

	fonte são aceites como estando corretos.	
Business rules conformance	O grau que indica a medida em que os dados estão em conformidade com as regras de negócio.	(English, 1999)
Clarity of definition	O grau que indica a medida em que se verifica que a definição dos dados está efetuada de uma forma clara.	(Loshin, 2001)
Completeness	O grau que indica a medida em que os dados podem atender às necessidades de informação de um utilizador, com uma amplitude, profundidade e alcance suficientes para a tarefa em questão.	(Calero et al., 2008; Moraga et al., 2009)
Compliance	O grau que indica a medida em que os dados possuem atributos que vão de encontro aos padrões, convenções ou regulamentos em vigor e regras similares relativas à qualidade dos dados num contexto específico de uso.	(Moraga et al., 2009)
Concise Representation	O grau que indica a medida em que os dados estão representados de uma forma concisa, sem elementos supérfluos e que permitam a deteção de descrições incorretas.	(Calero et al., 2008; Moraga et al., 2009)
Confidentiality	O grau que indica a medida em que os dados possuem atributos que garantem que só são acessíveis e interpretáveis por utilizadores autorizados num contexto específico de uso.	(Moraga et al., 2009)
Consistency	O grau que indica a medida em que um conjunto de dados possui dados que não violam as regras semânticas que foram definidas, que possuem atributos que estão livres de contradições, que são coerentes com outros dados e que são apresentados no mesmo formato num contexto específico de uso.	(Batini et al., 2009; Moraga et al., 2009)
Consistent representation	O grau que indica a medida em que os dados são sempre apresentados no mesmo formato, que são compatíveis com dados anteriores e consistentes com outras fontes.	(Calero et al., 2008; Moraga et al., 2009)
Correct Interpretation	O grau que indica a medida em que é efetuada uma interpretação correta dos dados, na sua apresentação.	(Loshin, 2001)
Correctness	O grau que indica a medida em que os dados estão	(Batini et al., 2006)

	corretos, tendo em conta os requisitos ou o modelo.	
Cost	O grau que indica a medida dos custos da falta de qualidade dos dados e os custos da melhoria dessa qualidade.	(Batini et al., 2006)
Credibility	O grau que indica a medida em que os dados possuem atributos considerados verdadeiros, corretos e credíveis pelos utilizadores.	(Moraga et al., 2009)
Currency	O grau que indica a medida em os dados que se encontram no portal não estão obsoletos, ou seja, se os dados se encontram atualizados, apesar de possíveis discrepâncias causadas por alterações temporárias ao valor correto.	(Batini et al., 2009; Moraga et al., 2009)
Currentness	O grau que indica a medida em que os dados possuem atributos que não se encontram obsoletos e estão atualizados num contexto específico de uso.	(Moraga et al., 2009)
Customer Support	O grau que indica a medida em que o portal Web onde se encontram os dados fornece suporte on-line através de texto, email, telefone, etc.	(Calero et al., 2008; Moraga et al., 2009)
Documentation	Grau que indica a quantidade e utilidade dos documentos com informação dos metadados.	(Calero et al., 2008; Moraga et al., 2009)
Duplicates	O grau que indica a medida em que os dados no portal contêm dados duplicados.	(Calero et al., 2008)
Ease of manipulation	O grau de facilidade de manipulação dos dados.	(Pipino et al., 2002)
Ease of operation	O grau que indica a medida em que os dados são facilmente geridos, podem ser aplicados em diferentes tarefas e podem ser manipulados (ou seja, atualizados, movidos, agregados, etc.)	(Calero et al., 2008; Moraga et al., 2009)
Effectiveness	O grau que indica a medida em que até que ponto são utilizadas técnicas analíticas adequadas nos dados que estão no portal.	(Moraga et al., 2009)
Efficiency	O grau que indica a medida em que os dados do portal possuem atributos que podem ser processados e fornecem os níveis de desempenho esperados, através da utilização de tipos e quantidades adequadas de recursos.	(Moraga et al., 2009)

Expiration	O grau que indica a data conhecida de validade dos dados.	(Moraga et al., 2009)
Flexibility	O grau que indica a medida em que os dados são expansíveis, adaptáveis e facilmente aplicados a outras necessidades.	(Calero et al., 2008; Moraga et al., 2009)
Freedom from errors	O grau que indica o nível de ausência de erros dos dados.	(Lee et al., 2002)
Interactive	O grau que indica a medida em que a forma como os dados são acedidos ou recuperados pode ser adaptada às preferências pessoais de alguém através de elementos interativos.	(Calero et al., 2008; Moraga et al., 2009)
Interpretability	O grau que indica a medida em que os dados do portal encontram-se com um idioma e unidades que são apropriadas para a capacidade do utilizador.	(Calero et al., 2008; Moraga et al., 2009)
Meaningfulness	O grau que indica a medição da significância dos dados.	(Su & Jin, 2004)
Metadata	O grau que indica a medição da informação sobre os metadados dos dados.	(Loshin, 2001)
Novelty	O grau que indica a medida em que os dados obtidos do portal influenciam o conhecimento e as novas decisões.	(Calero et al., 2008; Moraga et al., 2009)
Null Values	O grau que indica a quantidade de dados com valores a *null*.	(Loshin, 2001)
Objectivity	O grau que indica a medida em que os dados são imparciais, sem preconceitos e não tendenciosos.	(Calero et al., 2008; Moraga et al., 2009)
Organization	A organização, configurações visuais ou características tipográficas (cor, texto, fonte de letra, imagens, etc.) e as combinações desses vários componentes.	(Calero et al., 2008; Moraga et al., 2009)
Portability	O grau que indica em que medida os dados possuem atributos que os permitem ser instalados, substituídos ou movidos de um sistema para outro, preservando a sua qualidade.	(Moraga et al., 2009)
Precision	O grau que indica em que medida os dados do portal possuem atributos que são exatos ou que oferecem discriminação de dados no portal e como ajudam os utilizadores a encontrar resultados relevantes e a	(Moraga et al., 2009)

	evitar os resultados irrelevantes.	
Privacy	Indicador da política de privacidade dos dados.	(Loshin, 2001)
Readability	A medida que indica que o texto é legível e é apresentado de uma forma fácil de ler.	(Moraga et al., 2009)
Recoverability	O grau que indica a medida em que os dados possuem atributos que os permitem manter e preservar um nível especificado de operações e qualidade, mesmo em caso de falha.	(Moraga et al., 2009)
Relevancy	O grau que indica em que medida os dados são aplicáveis, úteis e proveitosos para as necessidades dos utilizadores, na tarefa em questão.	(Bizer & Cyganiak, 2009; Calero et al., 2008; English, 1999; Long & Seko, 2002; Loshin, 2001; Moraga et al., 2009; Pipino et al., 2002; Su & Jin, 2004)
Reliability	A medida que indica que os utilizadores podem confiar nos dados, nas suas fontes e na perceção do utilizador final sobre o funcionamento técnico adequado do portal onde se encontram os dados.	(Calero et al., 2008; Moraga et al., 2009; Su & Jin, 2004)
Reputation	A medida que indica que os dados são confiáveis ou altamente considerados em termos da sua fonte ou conteúdo, e a informação é altamente tida em consideração em termos da sua origem ou conteúdo.	(Batini et al., 2006; Calero et al., 2008; Lee et al., 2002; Moraga et al., 2009; Pipino et al., 2002; Su & Jin, 2004; Wang, 1998)
Response time	A quantidade de tempo decorrido até que a resposta chega completa ao utilizador.	(Calero et al., 2008; Jeusfeld et al., 1998)
Scope	O grau que indica a medida em que os dados têm uma amplitude, profundidade e alcance suficientes para a tarefa em questão.	(Moraga et al., 2009)
Security	O grau que indica em que medida são passadas informações de forma privada do utilizador para a fonte da informação e vice-versa.	(Calero et al., 2008; Eppler & Helfert, 2004; Jeusfeld et al., 1998; Lee et al., 2002; Loshin, 2001; Pipino et al., 2002;

		Wang, 1998)
Specialization	Grau de especificidade dos dados, que deve incorporar todos os detalhes que podem ser vistos pelos utilizadores.	(Calero et al., 2008; Moraga et al., 2009)
Timeliness	O grau que indica em que medida os dados mudaram dentro dos limites de tempo especificados pela organização de destino, se os dados são transitórios ou estáveis e se os dados estão suficientemente atualizados para a tarefa em questão.	(Batini et al., 2009; Calero et al., 2008; Moraga et al., 2009)
Timeliness (Currency, Volatility)	Este grau contém dois componentes: a idade e a volatilidade. A idade é uma medida que indica a idade da informação, com base no tempo decorrido deste que os dados foram gravados. A volatilidade é uma medida de instabilidade da informação, ou seja, a frequência de alteração dos valores dos atributos das entidades.	(Batini et al., 2009)
Timeliness Comparability	Medida que indica a comparação temporal dos dados.	(Long & Seko, 2002)
Traceability	O grau que indica em que medida os dados estão bem documentados, verificáveis, facilmente atribuídos a uma fonte e fornecem uma via de auditoria de acesso aos dados e a qualquer alteração que tenha sido efetuada nos dados.	(Calero et al., 2008; Moraga et al., 2009)
Unambiguity	O grau que indica em que medida existem informações insuficientes, ou seja, os dados podem ser interpretados por mais de uma maneira.	(Su & Jin, 2004; Wand & Wang, 1996)
Understandability	O grau que indica em que medida os dados possuem atributos que os permitem ser lidos, que são claros, sem ambiguidade, facilmente compreensíveis, bem interpretados pelos utilizadores e que são expressos em linguagens, símbolos e unidades apropriadas num contexto específico de utilização.	(Calero et al., 2008; Moraga et al., 2009)
Uniqueness	O grau que indica o número de dados duplicados.	(Batini et al., 2009)
Usability	A usabilidade é definida pelo estado da acessibilidade, documentação e nível de interpretação dos dados.	(Long & Seko, 2002)

Usefulness	A extensão da avaliação do utilizador tanto da probabilidade de a informação melhorar as suas decisões como da sua satisfação com a utilidade do conteúdo, no que toca ao uso de linguagem apropriada e à utilidade da informação, de acordo com as necessidades da audiência a quem é dirigida.	(Moraga et al., 2009)
Validity	O grau que indica em que medida os utilizadores podem julgar e compreender os dados entregues pelo portal.	(Calero et al., 2008; Moraga et al., 2009)
Value added	O grau que indica a medida em que os dados são benéficos e proporcionam vantagens a partir da sua utilização.	(Calero et al., 2008; Moraga et al., 2009)
Verifiability	A extensão das referências a fontes originais.	(Moraga et al., 2009)

Tabela 7.2 – Tabela com o estudo do tipo de dados dos conjuntos de dados que se encontram disponibilizados no Portal de Dados Abertos da cidade de Lisboa

Grupo	Designação do Conjunto de Dados	Extensão	Dados Estruturados	Dados Semi-Estruturados	Dados Não Estruturados
Administração Pública e Justiça	Administração Central - Finanças	GeoJSON		X	
	CM Lisboa - Paços do Concelho				
	CM Lisboa - Empresas Municipais				
	CM Lisboa - Atendimento				
	CM Lisboa - Assembleia Municipal				
	CM Lisboa - Apoio à Juventude				
	Juntas de Freguesia				
	CM Lisboa - Animais				
	Administração Central - Ministérios	SHP		X	
	Administração Central - IMTT				
	CM Lisboa - Centros de Recepção de Resíduos				
Ambiente	Áreas de Gestão de Remoção	GeoJSON		X	

	Vidrões				
	Sondagens Geológicas				
	Postos de Limpeza				
	Parques Recreativos				
	Miradouros				
	Locais de recepção - Resíduos de Equipamentos Eléctricos e Electrónicos				
	Locais de Recepção - Óleos Alimentares Usados				
	Jardins - Parques Urbanos				
	Geomonumentos				
	Hortas Urbanas				
	Grandes Parques e Jardins de Lisboa				
	Espaços Verdes				
	Elementos de Água				
	Ecoponto				
	Ecoilhas Subterrâneas				
	Eco-ilhas				

	Corredor Verde				
	Circuitos de Recolha de Resíduos Urbanos				
	Centro de Recepção de Papel				
	Carta de Tipo de Solos				
	Arvoredo				
	CM Lisboa - Animais				
	CM Lisboa - Centros de Recepção de Resíduos	SHP		X	
	Modelo Batimétrico do Porto de Lisboa				
	Avisos Locais - Porto de Lisboa	PHP			X
	Previsão de Marés - Porto de Lisboa				
	Previsão do estado tempo para 5 dias para Lisboa	JSON		X	
	Carta do potencial solar de Lisboa	TIFF			X
	Carta de Declives				
	Modelo Digital de Terreno				
Cultura e Património	Topónimos Ibero-Americanos	GeoJSON		X	
	Vestígios Arqueológicos				
	Toponímia de Lisboa				

	Teatros				
	Residências Artísticas				
	Prémio Valmor (CML)				
	Património Mundial				
	Museus				
	Monumentos Nacionais				
	Intervenções Arqueológicas da CML				
	Imóveis, Monumentos e Conjuntos de Interesse Municipal				
	Imóveis em Vias de Classificação (Ministério da Cultura)				
	Imóveis em Vias de Classificação (CML)				
	Imóveis e Monumentos de Interesse Público				
	Geomonumentos				
	Galerias de Arte				
	Estatuária				
	Conjuntos de Interesse Público				
	Cinemas				
	Centros Culturais				

	Casas Regionais				
	Bibliotecas Arquivos e Centros de Documentação				
	Azulejaria				
	Auditórios - Anfiteatros				
	Arquitectura da Água				
	Arquitectura Religiosa				
	Arquitectura Premiada				
	Arquitectura Nobre				
	Arquitectura Militar				
	Arquitectura Industrial				
	Arquitectura Civil				
	Efemérides do quotidiano de Lisboa até 1920	CSV		X	
	Placas e lápides evocativas				
	Cartografia Histórica de Lisboa	GeoJSON, JSON		X	
	Medidas de Desempenho do Arquivo Municipal de Lisboa	CSV		X	
	Medidas de desempenho da Rede de Bibliotecas de Lisboa	CSV, PDF		X	X

	Inquérito à Satisfação com a Rede de Bibliotecas de Lisboa				
	Localização e identificação das Casas Religiosas de Lisboa existentes em 2015	GeoJSON		X	
	Orquestra Geração	XLS			X
	Galerias Municipais				
Educação	Outras Instituições de I&D e Fundações	GeoJSON	X		
	Laboratórios Associados				
	Faculdades, Escolas e Institutos				
	Escolas Públicas - Secundário				
	Escolas Públicas - Pré-Escolar				
	Escolas Públicas - 3º Ciclo				
	Escolas Públicas - 2º Ciclo				
	Escolas Públicas - 1º Ciclo				
	Escolas Profissionais				
	Escolas Privadas - Secundárias				
	Escolas Privadas - Pré-Escolar				
	Escolas Privadas - 2º e 3º Ciclo				

	Escolas Privadas - 1º Ciclo				
	Ensino Superior				
	Ensino Superior				
	Centros de Investigação e de Estudos				
	Equipamentos Escolares	XLSX			X
	Agrupamentos de Escolas de Lisboa				
	Beneficiários da Ação Social Escolar				
	Programa de Apoio à Natação Curricular				
	Orquestra Geração	XLS			
	Passaporte Escolar				
	População Escolar				
	Alimentação Escolar	XLSX			
Segurança e Socorro	Sistema de Videovigilância Bairro Alto	GeoJSON			X
	Protecção Civil				
	Polícia de Segurança Pública				
	Polícia Municipal				
	Bombeiros				

	Índice de Atropelamentos em Lisboa	TIFF			X
Desporto	Outras Instalações Desportivas (Dados 2009)	GeoJSON		X	
	Instalações Desportivas - Municipais e Concessões (Dados 2015)				
	Equipamentos de Fitness ao Ar Livre				
	Desporto Entidades Desportivas				
	Desporto Actividades Radicais				
	Eventos e Programas Desportivos previstos para Lisboa	XLSX, XLS			X
	Programa Desporto Mexe Comigo	XLS			
	Programa de Apoio à Natação Curricular	XLSX			
	Programa Clubes de Mar	XLS			
Gestão Urbana	Áreas suscetíveis de recurso ao Instrumento Financeiro para a Reabilitação e Revitalização Urbanas (IFRRU 2020)	GeoJSON		X	
	Repavimentações UCT				
	Publicidade em Painéis				
	Publicidade em MUPE (Mobiliário Urbano de Promoção Económica)				
	Processos de obras de edificação e demolição abrangidas pelo RJUE				

	Ocupações Temporárias de Espaço Público - Licenciadas				
	Ocupações Temporárias de Espaço Público - Histórico				
	Ocupaçoes Temporárias de Espaço Público Agendadas não Licenciadas				
	Limite de Concelho Oficial				
	Limite Unidades Territoriais - Macro				
	Intervenções Diversas				
	Instalações Sanitárias Públicas Automáticas				
	Freguesias-2012				
	Candeeiros de Iluminação Pública				
	Alvarás para obras de edificação e demolição emitidos ao abrigo do RJUE				
	Principais obras de promoção da SRU-OCIDENTAL	SHP		X	
	Principais Obras de Iniciativa Municipal				
	Principais obras de promoção da EMEL				
	Edifícios com Necessidade de Grandes Reparações	XLS			X
	Idade Média dos Edifícios				
	Edifícios Licenciados para Construção				

	Habitação Social				
	Áreas suscetíveis de recurso ao Instrumento Financeiro para a Reabilitação e Revitalização Urbanas (IFRRU 2020)				
Economia e Inovação	Construção e Reparação Naval	GeoJSON		X	
	Biotecnologia				
	Atividades Portuárias e Marítimas				
	Artes Performativas - Teatro, Dança e Música				
	Ambiente e Exploração Marítima				
	Turismo Náutico				
	Rádio e Televisão				
	Recenseamento Comercial 2010				
	Recenseamento Comercial 2009				
	Publicidade				
	Recenseamento Comercial 2008				
	Recenseamento Comercial 2007				
	Recenseamento Comercial 2006				
	Recenseamento Comercial 2005				
	Recenseamento Comercial 2004				

	Recenseamento Comercial 2002				
	Recenseamento Comercial 2000				
	Recenseamento Comercial 1998				
	Recenseamento Comercial 1996				
	Recenseamento Comercial 1995				
	Recenseamento Comercial 1993				
	Recenseamento Comercial 1991				
	Produção de Hardware				
	Produção de Fármacos				
	Prestação de Cuidados				
	Pesca e Derivados				
	Outras Instituições de I&D e Fundações				
	Náutica de Recreio e Património				
	Nanotecnologia				
	Música				
	Museus, Bibliotecas e Arquivos				
	Mercados				

	Mapa do Conhecimento - Incubadoras				
	Laboratórios Associados				
	Institutos				
	Instituições				
	Infraestruturas e Parques de Ciência e Tecnologia				
	Impressão e Reprodução Gráfica				
	I & D				
	I & D				
	Genómica				
	Fundações				
	Governação e Outras Entidades				
	Feiras				
	Faculdades, Escolas e Institutos				
	Espaços e Bairros Criativos				
	Espaços e Ambientes Criativos				
	Ensino de Atividades Criativas e Culturais				
	Ensino Superior				

	Ensino				
	Ensino				
	Edição (Livros, Jornais e Revistas)				
	Design				
	Defesa e Segurança				
	Cosmética				
	Comercialização de Hardware e Software e Serviços				
	Cinema e Video				
	Centros de Investigação e de Estudos				
	Centros de Investigação				
	Centros de Formação Profissional e Outros				
	Centros Comerciais				
	Associações				
	Arquitetura				
	Apoio e Financiamento à Inovação				
	Ganho Médio Mensal	JSON			
	Constituição e Dissolução de Empresas	XLS, JSON			X

	Volume de negócios das empresa	XLS			
	Índice de Polarização de Emprego				
	Taxa de Sobrevivência das Empresas				
	Volume de Negócios dos Estabelecimentos				
	Número de Estabelecimentos				
	Número de empresas				
Energia e Comunicações	Postos de Carregamento Mobi E	GeoJSON		X	
	Consumos Eléctricos Edifícios CML	XLS			X
	Instalações de Iluminação Pública	SHP		X	
	Traçado e Georeferenciação da Rede de Distribuição Elétrica				
	Carta do potencial solar de Lisboa	TIFF			X
Planeamento Urbano	Área de Reabilitação Urbana (ARU)	GeoJSON		X	
	Área Urbana de Génese Ilegal (AUGI)				
	Limite de Planos de Urbanização				
	Limite de Planos de Pormenor				
	Limite de Concelho Oficial				
	Limite de Concelho				

	Superfície das unidades territoriais por localização geográfica (NUTS - 2013)	JSON			
Habitação e Desenvolvimento Social	Contratos de Empreitadas de Obras Públicas e de Aquisição de Serviços	XLSX			X
	Fundo de Emergência Social de Lisboa - Agregados Familiares - Pedidos de apoio solicitados e concedidos por finalidade dos apoios				
	Fundo de Emergência Social de Lisboa - IPSS e outras entidades sem fins lucrativos				
	Serviço de Teleassistência (STA) - N.º de equipamentos de teleassistência instalados, ativos e n.º de beneficiários apoiados por freguesia				
	Apoio Financeiro ao abrigo do Regulamento de Atribuição de Apoios pelo Município de Lisboa (RAAML)				
	Programas de Apoio à Habitação				
	Beneficiários do RSI	JSON		X	
	Edifícios por Localização Geográfica e Tipo	XLSX			X
	Edifícios por Tipo de Utilização	XLS			
	Edifícios por Dimensão de pisos e Época de construção	XLSX			
	Alojamentos por Tipo de Alojamento Face à Forma de				

	Ocupação e Edifício				
	Taxa de Desemprego	XLS			
	Edifícios com Necessidade de Grandes Reparações				
	Encargos Médios com Habitação				
	Idade Média dos Edifícios				
	Alojamentos familiares clássicos e Regime de ocupação	XLSX			
	Edifícios Licenciados para Construção	XLS			
	Edifícios por Dimensão de Pisos e Materiais Usados na Construção	XLSX			
	Valor Médio das Rendas de Habitação Social	XLS			
	Habitação Social				
	Alojamentos Familiares Clássicos, por Época de Construção e Existência de Estacionamento ou Garagem	XLSX			
	Alojamentos Familiares Clássicos por Escalão de Divisões e Escalão de Área Útil				
	Alojamentos Familiares e Existência de Instalações				
	Alojamentos Familiares por Existência de Água Canalizada				
	Alojamentos Familiares com Existência de Instalação de Banho ou Duche				

	Programa BIP/ZIP (Bairros e Zonas de Intervenção Prioritária) - Projetos aprovados	XLS			
	Apoios Financeiros às Candidaturas ao Programa BIP/ZIP				
	Candidaturas a programas municipais de acesso à habitação e de apoio ao arrendamento				
	Valores das rendas de habitação social do Município de Lisboa				
	Dados Síntese do Património Gerido pela GEBALIS,EM	XLSX			
Informação Base e Cartografia	Topónimos	GeoJSON		X	
	Rede Viária (Escala entre 1/30000 e 1/20000)				
	Rede Viária (Escala >1/30000)				
	Rede Viária (Escala <1/20000)				
	Rede Ferroviária				
	Quarteirões				
	Limite de Concelho Oficial				
	Limite de Concelho				
	Freguesias-2012				
	Grandes Parques e Jardins de Lisboa				
	Plantas de Freguesia	PDF			X

	Número de Freguesias	XLS			
População	Saldo Migratório	JSON		X	
	Estimativas da população residente				
	Beneficiários do RSI				
	Taxa Bruta de Natalidade por 1000 habitantes				
	Índice de Envelhecimento				
	Taxa de Crescimento Migratório				
	Ganho Médio Mensal				
	População Residente por Sexo, Grupo Etário e Escalão de Dimensão Populacional	XLSX			X
	Taxa de Desemprego	XLS			
	Proporção da População Residente com Ensino Superior Completo				
	População em Lugares Urbanos				
	População Residente com 15 e Mais Anos de Idade	XLSX			
	População Empregada por Sexo, Sector de Atividade Económica e Situação na Profissão				
	Famílias Clássicas, por Dimensão e Tipo de Família (com Base na Estrutura Etária)				

	População Residente por Sexo, Idade e Escalão de Dimensão Populacional				
	População Residente com 10 e Mais Anos de Idade (Analfabetos)				
	População residente por sexo, grupo etário e nível de escolaridade				
	Famílias Institucionais, Condição Perante o Trabalho, Dimensão e Tipo				
	Famílias Clássicas, por Grupo Etário e Dimensão				
	Núcleos familiares, com base na idade dos filhos				
	Núcleos Familiares com Filhos com Menos de 6 anos de Idade				
	População Presente por Local de Residência e Sexo				
	População Residente Empregada ou Estudante por Sexo, Condição Perante o Trabalho e Local de Trabalho/Estudo				
	Famílias Clássicas, por Dimensão e Tipo de Família (com Base nos Núcleos Familiares)				
Outros Equipamentos e Serviços	Quiosques e Bancas	GeoJSON		X	
	Lavadouros				
	Instalações Sanitárias Públicas Automáticas				

	Instalações Sanitárias				
	Embaixadas				
	Conservatórias				
	Cemitérios				
	Balneários				
	Loja do Cidadão				
	Lojas Sociais	GeoJSON, XLS			X
	Ordens Profissionais	SHP			
	Locais de Culto				
	Estações de Correios				
	Centros de Emprego				
	Cartão do Cidadão				
Saúde	Hospitais Públicos	GeoJSON		X	
	Hospitais Privados				
	Hospitais Militares				
	Farmácias e Parafarmácias				

	Centros de Saúde				
	Clínicas	SHP			
Transparência	Catálogo de Dados Lisboa Aberta	XLSX			X
	Projetos vencedores referentes ao Orçamento Participativo de Lisboa				
	Balanço Social	XLS			
	Tabela de Taxas Municipais	XLSX, PDF			
	Tabela de Preços e Outras Receitas Municipais				
Turismo e Lazer	TukTuk - Estacionamentos	GeoJSON		X	
	Terminais de Cruzeiro				
	Parques de Merendas				
	Parques Recreativos				
	Parques Infantis				
	Miradouros				
	Jardins - Parques Urbanos				
	Equipamentos de Fitness ao Ar Livre				
	Elevadores e Ascensores				

	Docas de Recreio e Marinas				
	Casino				
	Alojamento				
	Preço Médio em Hotéis - Por Amostragem	XLS, XLSX			X
	Taxas de Ocupação Hoteleira - Por Amostragem	XLSX			
	Capacidade de Alojamento	XLS			
	Despesa Média de Turistas na Cidade	XLS, XLSX			
	Atividades de Animação Turística – Registo Nacional	JSON		X	
	Alojamento Local – Registo Nacional				
	Agências de Viagens e Turismo – Registo Nacional				
	Empreendimentos Turísticos – Registo Nacional				
	Taxas de Ocupação				
Mobilidade	Informação sobre Transportes Públicos da Cidade de Lisboa - Sul Fertagus	PDF, GTFS		X	
	TukTuk - Estacionamentos	GeoJSON			
	Terminais de Cruzeiro				
	Rede Viária (Escala entre 1/30000 e 1/20000)				
	Rede Viária (Escala >1/30000)				

	Rede Viária (Escala <1/20000)				
	Rede Ferroviária				
	Postos de Carregamento Mobi E				
	Painéis de Mensagem Variável				
	Localização de radares				
	Estações de Metro				
	Estações de Comboio				
	Estações Fluviais				
	Elevadores e Ascensores				
	Docas de Recreio e Marinas				
	Ciclovias da Cidade de Lisboa				
	Modelo Batimétrico do Porto de Lisboa	SHP			
	Avisos Locais - Porto de Lisboa	PHP			X
	Previsão de Marés - Porto de Lisboa				
	Informação sobre Transportes Públicos da Cidade de Lisboa - Metropolitano de Lisboa	PDF, GTFS		X	
	Informação sobre Transportes Públicos da Cidade de Lisboa - CP - Comboios de Portugal				

	Informação sobre Transportes Públicos da Cidade de Lisboa - Transportes Sul do Tejo				
	Informação sobre Transportes Públicos da Cidade de Lisboa - Fertagus				
	Informação sobre Transportes Públicos da Cidade de Lisboa - Transtejo				
	Informação sobre Transportes Públicos da Cidade de Lisboa - Carris				
	Informação sobre Transportes Públicos da Cidade de Lisboa - Soflusa				
	Informação sobre Transportes Públicos da Cidade de Lisboa - Rodoviária de Lisboa				
	EMEL - Sinalização EMEL	XLSX			
	EMEL - Parquímetros EMEL				
	EMEL - Lugares de estacionamento na via pública				
	EMEL - Parques de estacionamento na via pública				
	MAPPe - Mapa de Potencial Pedonal de Lisboa	TIFF			
	Declive Longitudinal da Rede Viária	GeoJSON		X	
	Índice de Atropelamentos em Lisboa	TIFF			

Tabela 7.3 – Tabela com o mapeamento entre os conjuntos de dados do Portal de Dados Abertos que foram estudados e as tabelas em SQL Server que foram criadas no âmbito deste projeto

Grupo de Dados	**Conjunto de Dados**	**Extensão**	**Tabelas – Base de Dados – 'LisbonOD'**
Administração Pública e Justiça	Administração Central - Finanças	GeoJSON	PublicAdmin.TaxOffice
Ambiente	Anomalia da média anual da temperatura média	CSV	Environment.AnomTempMediaAnual
Ambiente	Anomalia da média anual da temperatura máxima	CSV	Environment.AnomTempMaxAnual
Ambiente	Anomalia da média anual da precipitação	CSV	Environment.AnomPreciMedAnual
Ambiente	Anomalia da média anual da temperatura mínima	CSV	Environment.AnomTempMinAnual
Ambiente	Vidrões	GeoJSON	Environment.GlassWaste
Cultura e Património	Cinemas	GeoJSON	Culture.Cinema
Cultura e Património	Conjuntos de Interesse Público	GeoJSON	Culture.PublicInterestSet
Cultura e Património	Teatros	GeoJSON	Culture.Theater
Educação	Agrupamentos de Escolas de Lisboa	XLSX	Education.SchoolsGroup Education.SchoolsNonGroup
Educação	Faculdades, Escolas e Institutos	GeoJSON	Education.HighEduInstitutions
Educação	Escolas Públicas - Secundário	GeoJSON	Education.PublicSecSchool
Segurança e Socorro	Polícia de Segurança Pública	GeoJSON	SecurityAid.PoliceStation

Segurança e Socorro	Bombeiros	GeoJSON	SecurityAid.FireDep
Desporto	Desporto Entidades Desportivas	GeoJSON	Sport.SportEntity
Desporto	Instalações Desportivas - Municipais e Concessões (Dados 2015)	GeoJSON	Sport.SportFacility2015
Gestão Urbana	Publicidade em Painéis	GeoJSON	UrbanMan.AdvertisingPanel
Gestão Urbana	Candeeiros de Iluminação Pública	GeoJSON	UrbanMan.LightingFixture
Economia e Inovação	Museus, Bibliotecas e Arquivos	GeoJSON	EconInov.MuseumLibrArch
Economia e Inovação	Centros Comerciais	GeoJSON	EconInov.ShoppingCenter
Energia e Comunicações	Postos de Carregamento Mobi E	GeoJSON	EnergyCom.MobiE
Habitação e Desenvolvimento Social	Edifícios por Localização Geográfica e Tipo	XLSX	HousingDev.BuildingsGeoLocType
Habitação e Desenvolvimento Social	Alojamentos por Tipo de Alojamento Face à Forma de Ocupação e Edifício	XLSX	HousingDev.AccommodationByTypeOfAccomm
Habitação e Desenvolvimento Social	Edifícios por Tipo de Utilização	XLSX	HousingDev.BuildingsTypeOfUse
Informação Base e Cartografia	Topónimos	GeoJSON	Cartography.Toponym
População	População Empregada por Sexo, Sector de Atividade Económica e Situação na Profissão	XLSX	PopulationDem.PopEmployedBy_Sex_EAS_Sit

População	População residente por sexo, grupo etário e nível de escolaridade	XLSX	PopulationDem.ResidentPopBy_Sex_AgeG_EducL
População	População Residente Empregada ou Estudante por Sexo, Condição Perante o Trabalho e Local de Trabalho/Estudo	XLSX	PopulationDem.ResidentPopBy_Sex_Empl_Student
Outros Equipamentos e Serviços	Loja do Cidadão	GeoJSON	ServiceEqui.CitizenShop
Outros Equipamentos e Serviços	Embaixadas	GeoJSON	ServiceEqui.Embassy
Outros Equipamentos e Serviços	Instalações Sanitárias	GeoJSON	ServiceEqui.Lavatory
Saúde	Hospitais Públicos	GeoJSON	Health.PublicHospital
Saúde	Farmácias e Parafarmácias	GeoJSON	Health.Drugstore
Saúde	Centros de Saúde	GeoJSON	Health.HealthCenter
Transparência	Projetos vencedores referentes ao Orçamento Participativo de Lisboa	XLSX	Transparency.ParticipatoryBudget
Turismo e Lazer	TukTuk - Estacionamentos	GeoJSON	Tourism.TukTukParkLot
Turismo e Lazer	Alojamento	GeoJSON	Tourism.HotelEstablishment
Turismo e Lazer	Jardins - Parques Urbanos	GeoJSON	Tourism.GardensAndParks
Mobilidade	Estações de Metro	GeoJSON	Mobility.UndergroundStation

Mobilidade	Estações de Comboio	GeoJSON	Mobility.RailwayStation
Mobilidade	Estações Fluviais	GeoJSON	Mobility.RiverPier

Tabela 7.4 – Tabela com a descrição de todos os campos que compõem a tabela 'Staging_Area.SA_Metric'

Chave	**Campo**	**Tipo**	**Null**	**Descrição**
PK	SK_Metric	Int	Not Null	Surrogate Key
	MetricID	Int	Not Null	Identificador da Métrica
	MetricName	Nvarchar(100)	Not Null	Designação da Métrica
	MetricDescription	Nvarchar(MAX)	Not Null	Descrição da Métrica
	MetricFormula	Nvarchar(300)	Null	Formula da Métrica
	Active	Bit	Not Null	Estado da Métrica (1 – Ativa, 0 – Inativa)
	DQ_DimensionName	Nvarchar(100)	Not Null	Designação da Dimensão a que pertence a Métrica
	DQ_DimensionDescription	Nvarchar(MAX)	Null	Descrição da Dimensão a que pertence a Métrica
	MethodologyName	Nvarchar(100)	Not Null	Designação da Metodologia ao qual pertence a Dimensão
	MethodologyDescription	Nvarchar(200)	Null	Descrição da Metodologia ao qual pertence a Dimensão
	Period_ID	Nchar(10)	Not Null	Identificador do Período

Tabela 7.5 - Tabela com a descrição de todos os campos que compõem a tabela 'Staging_Area.SA_TableOD'

Chave	**Campo**	**Tipo**	**Null**	**Descrição**
PK	SK_TableOD	Int	Not Null	Surrogate Key
	TableID	Int	Null	Identificador da Tabela
	TableSchema	Nvarchar(100)	Null	Designação do *Schema* da Tabela
	TableName	Nvarchar(200)	Null	Designação da Tabela
	Source	Nvarchar(200)	Null	Fonte de informação da Tabela
	Author	Nvarchar(200)	Null	Autor da fonte de informação da Tabela
	LastUpdate	Date	Null	Data da última atualização da Tabela
	CreateDate	Date	Null	Data da criação da Tabela
	Actualization	Nvarchar(30)	Null	Indicação do Período de Atualização
	ActualizationMonths	Int	Null	Período de Atualização em número de meses
	LastDataActualization	Date	Null	Data da última atualização dos dados da Tabela
	Language	Nvarchar(50)	Null	Linguagem contida na Tabela
	Publisher	Nvarchar(100)	Null	Indicação do divulgador da Tabela
	AttributeID	Int	Null	Identificador do atributo da Tabela

	AttributeName	Nvarchar(200)	Null	Designação do atributo da Tabela
	OrdinalPosition	Int	Null	Posição ordinal do atributo na Tabela
	DataGroupID	Int	Null	Identificador do grupo de dados da Tabela
	DataGroupName	Varchar(100)	Null	Designação do grupo de dados da Tabela
	DataBaseName	Nvarchar(200)	Null	Designação da base de dados onde se encontra a Tabela
	Period_ID	Nchar(10)	Not Null	Identificador do Período

Tabela 7.6 - Tabela com a descrição de todos os campos que compõem a tabela 'Staging_Area.Measurement'

Chave	**Campo**	**Tipo**	**Null**	**Descrição**
PK	SK_Measurement	Int	Not Null	Surrogate Key
	MetricID	Int	Not Null	Identificador da Métrica
	TableOD_ID	Int	Null	Identificador da Tabela
	MeasureValue	Decimal (10,2)	Not Null	Valor da métrica aplicada a um conjunto de dados ou atributo
	MeasurementUnit	Char(10)	Null	Unidade de valor da métrica
	Period_ID	Nchar(10)	Not Null	Identificador do Período

Tabela 7.7 - Tabela com a descrição de todos os campos que compõem a tabela 'Dimensions.D_Date'

Chave	**Campo**	**Tipo**	**Null**	**Descrição**
PK	Date_ID	Int	Not Null	Identificador da Data
	Date_day	Date	Null	Data com o formato 'DD/MM/AAAA'
	Full_date	Nvarchar(50)	Null	Data com o formato 'DD MM AAAA', com o mês escrito por extenso
	Day_number	Int	Null	Dia
	Day_name	Nvarchar(50)	Null	Nome do Dia
	Day_name_short	Nvarchar(50)	Null	Nome do Dia Resumido
	Week_day_FLG	Nvarchar(50)	Null	Indicação se é dia útil
	Month_number	Int	Null	Mês
	Month_name	Nvarchar(50)	Null	Nome do Mês
	Month_name_short	Nvarchar(50)	Null	Nome do Mês Resumido
	Trimester_number	Int	Null	Trimestre
	Trimester_name	Nvarchar(50)	Null	Nome do Trimestre
	Semester_number	Int	Null	Semestre
	Semester_name	Nvarchar(50)	Null	Nome do Semestre
	Year	Int	Null	Ano

Tabela 7.8 - Tabela com a descrição de todos os campos que compõem a tabela 'Dimensions.D_Metric'

Chave	**Campo**	**Tipo**	**Null**	**Descrição**
PK	MetricID	Int	Not Null	Surrogate Key
	Business_MetricID	Int	Not Null	Identificador da métrica
	MetricName	Nvarchar(100)	Not Null	Designação da Métrica
	MetricDescription	Nvarchar(MAX)	Not Null	Descrição da Métrica
	MetricFormula	Nvarchar(300)	Null	Formula da Métrica
	Active		Not Null	Estado da Métrica (1 – Ativa, 0 – Inativa)
	DQ_DimensionName	Nvarchar(100)	Not Null	Designação da Dimensão a que pertence a Métrica
	DQ_DimensionDescription	Nvarchar(MAX)	Null	Descrição da Dimensão a que pertence a Métrica
	MethodologyName	Nvarchar(100)	Not Null	Designação da Metodologia ao qual pertence a Dimensão
	MethodologyDescription	Nvarchar(200)	Null	Descrição da Metodologia ao qual pertence a Dimensão
	Period_ID	Nchar(10)	Not Null	Identificador do Período

Tabela 7.9 - Tabela com a descrição de todos os campos que compõem a tabela 'Dimensions.D_TableOD'

Chave	**Campo**	**Tipo**	**Null**	**Descrição**
PK	TableOD_ID	Int	Not Null	Surrogate Key
	TableID	Int	Null	Identificador da Tabela
	TableSchema	Nvarchar(100)	Null	Designação do *Schema* da Tabela
	TableName	Nvarchar(200)	Null	Designação da Tabela
	Source	Nvarchar(200)	Null	Fonte de informação da Tabela
	Author	Nvarchar(200)	Null	Autor da fonte de informação da Tabela
	LastUpdate	Date	Null	Data da última atualização da Tabela
	CreateDate	Date	Null	Data da criação da Tabela
	Actualization	Nvarchar(30)	Null	Indicação do Período de Atualização
	ActualizationMonths	Int	Null	Período de Atualização em número de meses
	LastDataActualization	Date	Null	Data da última atualização dos dados da Tabela
	Language	Nvarchar(50)	Null	Linguagem contida na Tabela
	Publisher	Nvarchar(100)	Null	Indicação do divulgador da Tabela
	AttributeID	Int	Null	Identificador do atributo da Tabela
	AttributeName	Nvarchar(200)	Null	Designação do atributo da Tabela

	OrdinalPosition	Int	Null	Posição ordinal do atributo na Tabela
	DataGroupID	Int	Null	Identificador do grupo de dados da Tabela
	DataGroupName	Varchar(100)	Null	Designação do grupo de dados da Tabela
	DataBaseName	Nvarchar(200)	Null	Designação da base de dados onde se encontra a Tabela
	Start_Date	Date	Not Null	Data de entrada
	End_Date	Date	Null	Data de saída
	Period_ID	Nchar(10)	Not Null	Identificador do Período

Tabela 7.10 - Tabela com a descrição de todos os campos que compõem a tabela 'Facts.F_Measure'

Chave	**Campo**	**Tipo**	**Null**	**Descrição**
PK	MeasureID	Int	Not Null	Identificador da Medição
	MetricID	Int	Not Null	Identificador da Métrica
	TableOD_ID	Int	Null	Identificador da Tabela
	MeasureValue	Decimal (10,2)	Not Null	Valor da métrica aplicada a um conjunto de dados ou atributo
	MeasurementUnit	Char(10)	Null	Unidade de valor da métrica
	Period_ID	Nchar(10)	Not Null	Identificador do Período